THE ANNALS
of
MATHEMATICAL STATISTICS

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Volume VIII, Number 3
September, 1937

PUBLISHED QUARTERLY
ANN ARBOR, MICHIGAN

Four Dollars per annum

Back numbers available at the following prices:
Vols. I–IV $5 each. Single numbers $1.50
Vol. V to date $4 each. Single numbers $1.25

Made in United States of America

Address: ANNALS OF MATHEMATICAL STATISTICS
Post Office Box 171, Ann Arbor, Michigan

Office of the Institute of Mathematical Statistics:
Secretary: ALLEN T. CRAIG, University of Iowa
Iowa City, Iowa

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD.

ON CERTAIN DISTRIBUTIONS DERIVED FROM THE MULTINOMIAL DISTRIBUTION[1]

By Solomon Kullback

1. Introduction. With the multinomial distribution as a background, there may be derived a number of distributions which are of interest in certain practical applications. Several of these distributions are here presented and the theory is illustrated by specific examples.

2. Preliminary data. In the discussion of the distributions to be considered there are needed certain factorial sums whose values are now to be derived. In the following discussion only positive integral values (including zero) are to be considered.

There is desired the value, in terms of N, n, r, of

$$f_r(n, N) = \sum \frac{N!}{x_1!\, x_2! \cdots x_n!} \tag{2.1}$$

where the summation is for all values of $x_1, x_2, \cdots, x_n$ such that $x_1 + x_2 + \cdots + x_n = N$ and no x is equal to r.

Let us first consider the case for $r = 0$; i.e., we desire a value for the sum in (2.1) for all values of $x_1, x_2, \cdots, x_n$ such that $x_1 + x_2 + \cdots + x_n = N$ and no x is equal to zero. By the multinomial theorem, we have that[2]

$$(a_1 + a_2 + \cdots + a_n)^N = \sum \frac{N!}{x_1!\, x_2! \cdots x_n!} a_1^{x_1} a_2^{x_2} \cdots a_n^{x_n} \tag{2.2}$$

where the summation is for all values of $x_1, x_2, \cdots, x_n$ such that $x_1 + x_2 + \cdots + x_n = N$. If $a_1 = a_2 = \cdots = a_n = 1$, then

$$n^N = \sum \frac{N!}{x_1!\, x_2! \cdots x_n!}, \qquad x_1 + x_2 + \cdots + x_n = N. \tag{2.3}$$

The sum in (2.3) may however be rearranged into the sum of a number of terms as follows:

$$\tag{2.4} \begin{cases} \displaystyle\sum \frac{N!}{x_1!\, x_2! \cdots x_n!}, & x_1 + x_2 + \cdots + x_n = N, \quad \text{no } x = 0; \\ \displaystyle n \sum \frac{N!}{x_1!\, x_2! \cdots x_{n-1}!}, & x_1 + x_2 + \cdots + x_{n-1} = N, \quad \text{no } x = 0; \\ \displaystyle \frac{n(n-1)}{2} \sum \frac{N!}{x_1!\, x_2! \cdots x_{n-2}!}, & x_1 + x_2 + \cdots + x_{n-2} = N, \quad \text{no } x = 0; \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \\ \displaystyle \binom{n}{r} \sum \frac{N!}{x_1!\, x_2! \cdots x_{n-r}!}, & x_1 + x_2 + \cdots + x_{n-r} = N, \quad \text{no } x = 0. \end{cases}$$

[1] Presented to the Institute of Mathematical Statistics January 2, 1936.

[2] H. S. Hall & S. R. Knight, *Higher Algebra*, MacMillan & Co., 4th Ed. (1924), Chap. 15.

Thus we may rewrite (2.3) as

$$n^N = f_0(n, N) + nf_0(n-1, N) \tag{2.5}$$
$$+ \frac{n(n-1)}{2!} f_0(n-2, N) + \cdots + \binom{n}{r} f_0(n-r, N) + \cdots$$

Replacing n by $n - 1$ in (2.5) there is obtained

$$(n-1)^N = f_0(n-1, N) \tag{2.6}$$
$$+ (n-1)f_0(n-2, N) + \cdots + \binom{n-1}{r} f_0(n-r-1, N) + \cdots$$

Multiplying (2.6) by n and subtracting the result from (2.5), there is obtained

$$n^N - n(n-1)^N = f_0(n, N) \tag{2.7}$$
$$- \frac{n(n-1)}{2!} f_0(n-2, N) - \cdots - r\binom{n}{r+1} f_0(n-r-1, N) - \cdots$$

Replacing n by $n - 2$ in (2.5) there is obtained

$$(n-2)^N = f_0(n-2, N) \tag{2.8}$$
$$+ (n-2)f_0(n-3, N) + \cdots + \binom{n-2}{r-1} f_0(n-r-1, N) + \cdots$$

Multiplying (2.8) by $n(n - 1)/2$ and adding the result to (2.8), there is obtained

$$n^N - n(n-1)^N + \frac{n(n-1)}{2!}(n-2)^N = f_0(n, N) + \frac{n(n-1)(n-2)}{3!} \tag{2.9}$$
$$f_0(n-3, N) + \cdots + \frac{r(r-1)}{2!}\binom{n}{r+1} f_0(n-r-1, N) + \cdots$$

Continuing this process, there is finally obtained the result that

$$f_0(n, N) = n^N - n(n-1)^N + \frac{n(n-1)}{2!}(n-2)^N - \cdots \pm n \cdot 1^N \tag{2.10}$$

It may be shown[3] that the right side of (2.10) is $\Delta^n x^N$ for $x = 0$. The author has elsewhere obtained (2.10), but by a special procedure not applicable to the general case.[4]

We may readily verify (2.10) for example, for $n = 3$, $N = 5$. If $x_1 + x_2 + x_3 = 5$ and no $x = 0$, then the sets of solutions are (3,1,1), (1,3,1), (1,1,3), (2,2,1), (2,1,2), (1,2,2), and $f_0(3,5) = 3 \cdot \frac{5!}{3!\,1!\,1!} + 3 \frac{5!}{2!\,2!\,1!} = 150$. From (2.10) there is obtained $f_0(3,5) = 3^5 - 3.2^5 + 3.2/2 = 150$.

---

[3] E. T. Whittaker & G. Robinson, *The Calculus of Observations*, Blackie & Son Ltd. (1924), p. 7.

[4] S. Kullback, "On the Bernoulli Distribution," *Bull. Am. Math. Soc.*, December, 1935.

For the general case, we return again to (2.3) and rearrange the right side into the sum of a number of terms as follows:

$$(2.11)\quad \begin{cases} \sum \dfrac{N!}{x_1!\,x_2!\cdots x_n!}, & x_1 + x_2 + \cdots + x_n = N, \quad \text{no } x = r; \\ \dfrac{n}{r!}\sum \dfrac{N!}{x_1!\,x_2!\cdots x_{n-1}!}, & x_1 + x_2 + \cdots + x_{n-1} = N - r, \quad \text{no } x = r; \\ \dfrac{n(n-1)}{2!\,(r!)^2}\sum \dfrac{N!}{x_1!\,x_2!\cdots x_{n-2}!}, & x_1 + x_2 + \cdots + x_{n-2} = N - 2r, \quad \text{no } x = r; \\ \cdots\cdots & \cdots\cdots \\ \dbinom{n}{k}\left(\dfrac{1}{r!}\right)^k \sum \dfrac{N!}{x_1!\,x_2!\cdots x_{n-k}}, & x_1 + x_2 + \cdots + x_{n-k} = N - kr, \quad \text{no } x = r. \end{cases}$$

Thus we may rewrite (2.3) as

$$(2.12)\quad \begin{aligned} n^N = f_r(n, N) &+ \frac{nN^{(r)}}{r!} f_r(n-1, N-r) \\ &+ \frac{n(n-1)N^{(2r)}}{2!\,(r!)^2} f_r(n-2, N-2r) + \cdots \end{aligned}$$

where $N^{(k)} = N(N-1)(N-2)\cdots(N-k+1)$.

Replacing n by $n-1$ and N by $N-r$ in (2.12) there is obtained

$$(2.13)\quad \begin{aligned} (n-1)^{N-r} = f_r(n-1, N-r) &\\ &+ \frac{(n-1)(N-r)^{(r)}}{r!} f_r(n-2, N-2r) + \cdots \end{aligned}$$

Multiplying (2.13) by $\dfrac{nN^{(r)}}{r!}$ and subtracting the result from (2.12), there is obtained

$$(2.14)\quad n^N - \frac{nN^{(r)}}{r!}(n-1)^{N-r} = f_r(n, N) - \frac{n(n-1)N^{(2r)}}{2!\,(r!)^2} f_r(n-2, N-2r) - \cdots$$

By continuing this process, in a manner similar to that used for the case $r = 0$ there is finally obtained

$$(2.15)\quad \begin{aligned} f_r(n, N) = n^N - \frac{nN^{(r)}}{r!}(n-1)^{N-r} &+ \frac{n(n-1)N^{(2r)}}{2!\,(r!)^2}(n-2)^{N-2r} \\ &- \binom{n}{3}\frac{N^{(3r)}}{(r!)^3}(n-3)^{N-3r} + \cdots \end{aligned}$$

By setting $r = 0$ in (2.15), there is of course obtained the value already found in (2.10).

We may readily verify (2.15) for example, for $n = 3$, $N = 5$, $r = 2$. If $x_1 + x_2 + x_3 = 5$ and no $x = 2$, then the sets of solutions are (5,0,0), (0,5,0),

(0,0,5), (4,1,0), (1,4,0), (1,0,4), (4,0,1), (0,1,4), (0,4,1), (3,1,1), (1,3,1), (1,1,3), and $f_2(3,5) = 3\cdot 5!/5! + 6\cdot 5!/4! + 3\cdot 5!/3! = 93$. From (2.15) there is obtained $f_2(3,5) = 3^5 - 3\cdot 5\cdot 4\cdot 2^3/2! + 3\cdot 2\cdot 5\cdot 4\cdot 3\cdot 2/2!(2!)^2 = 93$.

The same method of procedure may be applied to evaluate

$$f_{rs\cdots t}(n, N) = \sum \frac{N!}{x_1!\,x_2!\cdots x_n!}, \qquad x_1 + x_2 + \cdots + x_n = N, \quad \text{no } x = r, s, \cdots, \text{or } t. \tag{2.16}$$

Thus, there is derived the result that

$$\begin{aligned} f_{rs}(n, N) = n^N &- n\left(\frac{N^{(r)}(n-1)^{N-r}}{r!} + \frac{N^{(s)}(n-1)^{N-s}}{s!}\right) \\ &+ n(n-1)\left(\frac{N^{(2r)}(n-2)^{N-2r}}{2!\,(r!)^2} + \frac{N^{(r+s)}(n-2)^{N-r-s}}{(r!)\,(s!)}\right. \\ &\left. + \frac{N^{(2s)}(n-2)^{N-2s}}{2!\,(s!)^2}\right) - n(n-1)(n-2)\left(\frac{N^{(3r)}(n-3)^{N-3r}}{3!\,(r!)^3}\right. \\ &\left. + \frac{N^{(2r+s)}(n-3)^{N-2r-s}}{2!\,(r!)^2\,(s!)} + \frac{N^{(r+2s)}(n-3)^{N-r-2s}}{2!\,(r!)\,(s!)^2} + \frac{N^{(3s)}(n-3)^{N-3s}}{3!\,(s!)^3}\right) \end{aligned} \tag{2.17}$$

We may readily verify (2.17) for example, for $n = 3$, $N = 5$, $r = 0$, $s = 2$. If $x_1 + x_2 + x_3 = 5$ and no $x = 0$ or 2, then the sets of solutions are (3,1,1), (1,3,1), (1,1,3) and $f_{02}(3,5) = 3\cdot 5!/3! = 60$. From (2.17) there is obtained $f_{02}(3,5) = 3^5 - 3(2^5 + 5\cdot 4\cdot 2^3/2) + 3\cdot 2(1/2! + 5\cdot 4/2! + 5\cdot 4\cdot 3\cdot 2/(2!)^3) = 60$. It will be shown later (see section 8) that

$$\begin{aligned} f_r(n, N) = f_{rs}(n, N) &+ \frac{nN^{(s)}}{s!} f_{rs}(n-1, N-s) \\ &+ \frac{n(n-1)N^{(2s)}}{2!\,(s!)^2} f_{rs}(n-2, N-2s) + \cdots \end{aligned} \tag{2.18}$$

$$\begin{aligned} f_s(n, N) = f_{rs}(n, N) &+ \frac{nN^{(r)}}{r!} f_{rs}(n-1, N-r) \\ &+ \frac{n(n-1)N^{(2r)}}{2!\,(r!)^2} f_{rs}(n-2, N-2r) + \cdots \end{aligned} \tag{2.19}$$

From (2.18) and (2.19) there may be derived, by a method similar to that employed in deriving (2.15), that

$$\begin{aligned} f_{rs}(n, N) = f_r(n, N) &- \frac{nN^{(s)}}{s!} f_r(n-1, N-s) \\ &+ \frac{n(n-1)N^{(2s)}}{2!\,(s!)^2} f_r(n-2, N-2s) - \cdots \end{aligned} \tag{2.20}$$

This latter result also follows from (2.17 and (2.15).

Let us now consider the following generalization of (2.1). There is desired in terms of $N, n, r, a_1, a_2, \cdots, a_n$, the value of

$$F_r(n, N, a_1, a_2, \cdots, a_n) = \sum \frac{N!}{x_1!\, x_2! \cdots x_n!} a_1^{x_1} a_2^{x_2} \cdots a_n^{x_n} \tag{2.21}$$

where $a_1, a_2, \cdots, a_n$, are constants and the summation is for all values of $x_1, x_2, \cdots, x_n$ such that $x_1 + x_2 + \cdots + x_n = N$ and no $x = r$. The method of procedure is the same as that for the case already considered, viz when $a_1 = a_2 = \cdots = a_n = 1$.

The sum in (2.2) may be rearranged into the sum of a number of terms as follows:

$$(2.22)\quad \begin{cases} \sum \dfrac{N!}{x_1!\, x_2! \cdots x_n!} a_1^{x_1} a_2^{x_2} \cdots a_n^{x_n}, \quad x_1 + x_2 + \cdots + x_n = N, \quad \text{no } x = r; \\ \dfrac{a_1^r}{r!} \sum \dfrac{N!}{x_2! \cdots x_n!} a_2^{x_2} \cdots a_n^{x_n} + \cdots + \dfrac{a_n^r}{r!} \sum \dfrac{N!}{x_1! \cdots x_{n-1}!} a_1^{x_1} \cdots a_{n-1}^{x_{n-1}}, \\ \qquad\qquad x_1 + x_2 + \cdots + x_{n-1} = N - r, \text{ etc.}, \quad \text{no } x = r; \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{a_1^r \cdots a_k^r}{(r!)^k} \sum \dfrac{N!}{x_{k+1}! \cdots x_n!} a_{k+1}^{x_{k+1}} \cdots a_n^{x_n} + \cdots \\ \qquad + \dfrac{a_{n-k+1}^r \cdots a_n^r}{(r!)^k} \sum \dfrac{N!}{x_1! \cdots x_{n-k}!} a_1^{x_1} \cdots a_{n-k}^{x_{n-k}}, \\ \qquad\qquad x_1 + x_2 + \cdots + x_{n-k} = N - kr, \text{ etc.}, \quad \text{no } x = r; \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

For convenience, let us write

$$(2.23)\quad \begin{cases} A(n, N) = (a_1 + a_2 + \cdots + a_n)^N \\ A_i(n-1, N) = (a_1 + \cdots + a_{i-1} + a_{i+1} + \cdots + a_n)^N \\ A_{ij}(n-2, N) = (a_1 + \cdots + a_{i-1} + a_{i+1} + \cdots + a_{j-1} + a_{j+1} + \cdots + a_n)^N \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ G_r(n, N) = F_r(n, N, a_1, a_2, \cdots, a_n) \\ G_r(n-1, N, a_i) = F_r(n-1, N, a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n) \\ G_r(n-2, N, a_i, a_j) = F_r(n-2, N, a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_{j-1}, a_{j+1}, \cdots, a_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

so that (2.2) may be written as

$$(2.24)\quad \begin{aligned} A(n, N) &= G_r(n, N) + \frac{N^{(r)}}{r!} \sum_{i=1}^{n} a_i^r G_r(n-1, N-r, a_i) \\ &\quad + \frac{N^{(2r)}}{2!\,(r!)^2} \sum_{i,j=1}^{n} a_i^r a_j^r G_r(n-2, N-2r, a_i, a_j) + \cdots \qquad (i \neq j, \text{ etc.}) \end{aligned}$$

From (2.24), there are obtained n equations

$$A_i(n-1, N-r) = G_r(n-1, N-r, a_i) + \frac{(N-r)^{(r)}}{r!} \sum_{j=1}^{n} a_j^r G_r(n-2, N-2r, a_i, a_j) + \cdots \quad (i = 1, 2, \cdots, n, j \neq 1) \tag{2.25}$$

Multiplying (2.25) by $a_i^r N^{(r)}/r!$ and subtracting the result from (2.24), there is obtained

$$A(n, N) - \sum_{i=1}^{n} \frac{a_i^r N^{(r)}}{r!} A_i(n-1, N-r) = G_r(n, N) - \frac{N^{(2r)}}{2!\,(r!)^2} \sum_{i,j=1}^{n} a_i^r a_j^r G_r(n-2, N-2r, a_i, a_j) - \cdots \quad (i \neq j, \text{etc.}). \tag{2.26}$$

Continuing this procedure, there is finally obtained

$$G_r(n, N) = F_r(n, N, a_1, a_2, \cdots, a_n) = A(n, N) - \frac{N^{(r)}}{r!} \sum_{i=1}^{n} a_i^r A_i(n-1, N-r) + \frac{N^{(2r)}}{2!\,(r!)^2} \sum_{i,j=1}^{n} a_i^r a_j^r A_{ij}(n-2, N-2r) - \cdots \quad (i \neq j, \text{etc.}) \tag{2.27}$$

Similar results are obtainable for

$$G_{rs\cdots t} = F_{rs\cdots t}(n, N, a_1, a_2, \cdots, a_n) = \sum \frac{N!}{x_1!\, x_2! \cdots x_n!} a_1^{x_1} a_2^{x_2} \cdots a_n^{x_n} \tag{2.28}$$

where the summation is for all values of x_i such that $x_1 + x_2 + \cdots + x_n = N$, and no $x = r, s, \cdots$, or t.

Thus, it will be shown later (see section 8), that

$$G_r(n, N) = G_{rs}(n, N) + \frac{N^{(s)}}{s!} \sum_{i=1}^{n} a_i^s G_{rs}(n-1, N-s, a_i) + \frac{N^{(2s)}}{2!\,(s!)^2} \sum_{i,j=1}^{n} a_i^s a_j^s G_{rs}(n-2, N-2s, a_i, a_j) + \cdots \quad (i \neq j, \text{etc.}) \tag{2.29}$$

Corresponding to the derivation of (2.27), there is obtained from (2.29) the fact that

$$G_{rs}(n, N) = G_r(n, N) - \frac{N^{(s)}}{s!} \sum_{i=1}^{n} a_i^s G_r(n-1, N-s, a_i) + \frac{N^{(2s)}}{2!\,(s!)^2} \sum_{i,j=1}^{n} a_i^s a_j^s G_r(n-2, N-2s, a_i, a_j) - \cdots \quad (i \neq j, \text{etc.}) \tag{2.30}$$

3. **The problem to be studied.** Consider a trial in which one of n mutually exclusive events may occur, with the respective probabilities of occurrence

$p_1, p_2, \cdots, p_n$ where $p_1 + p_2 + \cdots + p_n = 1$. The probabilities of the various combinations of events which are possible in N trials are given by the terms of the expansion of $(p_1 + p_2 + \cdots + p_n)^N$.

In the N trials some of the possible events may not occur, others may occur one, twice, etc. It is desired to study the distribution of the number of events which do not occur; the distribution of the number of events which occur once each, etc. The simultaneous distributions of the events above described are also to be studied.

For example, the possible event may be the occurrence of a digit. A study of a sequence of random digits, in sets of ten, yielded the following three sample sets.

0	1	2	3	4	5	6	7	8	9
1	0	2	1	1	2	1	0	0	2
1	1	1	1	1	1	2	0	1	1
0	0	2	1	2	1	2	1	0	1

FIG. 1

In the first set three events do not occur, four occur once each, and three occur twice each. In the second set one event does not occur, eight events occur once each, and one event occurs twice; etc.

4. **Distribution of the number of events not occurring.** To obtain the distribution of the number of events which do not occur, there is applied to the expansion of $(p_1 + p_2 + \cdots + p_n)^N$ a procedure similar to that employed in section 2.

Thus, if π_{r0} represents the probability for r events not occurring, then

$$
(4.1)\quad \begin{cases}
\pi_{00} = \sum \dfrac{N!}{x_1!\, x_2! \cdots x_n!} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n}, \quad x_1 + x_2 + \cdots + x_n = N, \\
\hfill \text{no } x = 0; \\
\pi_{10} = \sum \dfrac{N!}{x_2! \cdots x_n!} p_2^{x_2} \cdots p_n^{x_n} + \cdots + \sum \dfrac{N!}{x_1! \cdots x_{n-1}!} p_1^{x_1} \cdots p_{n-1}^{x_{n-1}}, \\
\hfill x_1 + x_2 + \cdots + x_{n-1} = N, \text{etc.}, \quad \text{no } x = 0; \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\pi_{r0} = \sum \dfrac{N!}{x_{r+1}! \cdots x_n!} p_{r+1}^{x_{r+1}} \cdots p_n^{x_n} + \cdots + \sum \dfrac{N!}{x_1! \cdots x_{n-r}!} p_1^{x_1} \cdots p_{n-r}^{x_{n-r}}, \\
\hfill x_1 + x_2 + \cdots + x_{n-r} = N, \text{etc.}, \quad \text{no } x = 0; \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{cases}
$$

Employing (2.21), we may write (4.1) as

$$(4.2)\quad \begin{cases} \pi_{00} = F_0(n, N, p_1, p_2, \cdots, p_n) \\ \pi_{10} = F_0(n-1, N, p_2, \cdots, p_n) + \cdots + F_0(n-1, N, p_1, p_2, \cdots, p_{n-1}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \pi_{r0} = F_0(n-r, N, p_{r+1}, \cdots, p_n) + \cdots + F_0(n-r, N, p_1, \cdots, p_{n-r}) \end{cases}$$

Since $p_1 + p_2 + \cdots + p_n = 1$ there is found from (2.27) that

$$(4.3)\quad \begin{cases} \pi_{00} = 1 - \sum_{i=1}^{n} (1-p_i)^N + \frac{1}{2!} \sum_{i,j=1}^{n} (1-p_i-p_j)^N \\ \qquad\qquad - \frac{1}{3!} \sum_{i,j,k=1}^{n} (1-p_i-p_j-p_k)^N + \cdots \\ \pi_{10} = \sum_{i=1}^{n} (1-p_i)^N - \sum_{i,j=1}^{n} (1-p_i-p_j)^N \\ \qquad\qquad + \frac{1}{2!} \sum_{i,j,k=1}^{n} (1-p_i-p_j-p_k)^N - \cdots \\ \pi_{20} = \frac{1}{2!} \left\{ \sum_{i,j=1}^{n} (1-p_i-p_j)^N - \sum_{i,j,k=1}^{n} (1-p_i-p_j-p_k)^N + \cdots \right\} \\ \pi_{30} = \frac{1}{3!} \left\{ \sum_{i,j,k=1}^{n} (1-p_i-p_j-p_k)^N - \cdots \right\} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad (i \neq j, \text{ etc.}) \end{cases}$$

The factorial moments[5] of the distribution given by (4.3) are easily derived. The first factorial moment is given by $\sigma_1 = \pi_{10} + 2\pi_{20} + 3\pi_{30} + \cdots + r\pi_{r0} + \cdots$ and the summation of the proper terms in (4.3) yields

$$(4.4)\qquad \sigma_1 = \sum_{i=1}^{n} (1-p_i)^N$$

In general, the r-th factorial moment, given by $\sigma_r = \sum_{k=r}^{n} k(k-1) \cdots (k-r+1)\pi_{k0}$ is

$$(4.5)\qquad \sigma_r = \sum_{a,b,\cdots,r=1}^{n} (1-p_a-p_b-\cdots-p_r)^N, \qquad (a \neq b, \text{ etc.}).$$

Indeed, (4.3) illustrates the fact that, if $f(x)$ is the probability that a discontinuous variate takes the value x, then[6]

$$(4.6)\qquad f(x) = \frac{1}{x!} \sum_{k=0}^{n-x} (-1)^k \sigma_{x+k}/k!$$

[5] J. F. Steffensen, *Interpolation* (1927), p. 101.

[6] J. F. Steffensen, "Factorial Moments and Discontinuous Frequency Functions" *Skandinavisk Aktuarietidskrift*, Vol. VI (1923), pp. 73–89.

The moments about any constant of the distribution given by (4.3) may be derived from the factorial moments by the relation[7]

$$E(x-a)^r = (1 + \sigma_1\Delta + \sigma_2\Delta^2/2! + \cdots + \sigma_r\Delta^r/r!)\cdot\xi^r \qquad (\xi = -a) \tag{4.7}$$

where Δ is the difference operator of the calculus of finite differences, and ξ is replaced by $(-a)$ after the indicated operations have been performed.

Of special interest is the case when $p_1 = p_2 = \cdots = p_n = \dfrac{1}{n}$, for which (4.3) becomes

$$\left\{\begin{array}{l}
\pi_{00} = \left(\dfrac{1}{n}\right)^N f_0(n, N) = \left(\dfrac{1}{n}\right)^N \Delta^n 0^N \\
\pi_{10} = \left(\dfrac{1}{n}\right)^N n f_0(n-1, N) = \left(\dfrac{1}{n}\right)^N n\Delta^{n-1} 0^N \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\pi_{r0} = \left(\dfrac{1}{n}\right)^N \dbinom{n}{r} f_0(n-r, N) = \left(\dfrac{1}{n}\right)^N \dbinom{n}{r} \Delta^{n-r} 0^N \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{array}\right. \tag{4.8}$$

where $f_0(n, N)$ and $\Delta^n 0^N$ are as defined in section 2. The probabilities in (4.8) are the respective terms of the expansion of $\left(\dfrac{1}{n}\right)^N (1+\Delta)^n \cdot 0^N$.

For this case the r-th factorial moment becomes

$$\sigma_r = n(n-1)\cdots(n-r+1)\,(n-r)^N/n^N \tag{4.9}$$

There is presented an example of the distribution (4.8) for the case $n = N = 10$. It is found that[8]

$$\left\{\begin{array}{llll}
\Delta 0^{10} = 1 & & \Delta^6 0^{10} = 16435440 \\
\Delta^2 0^{10} = 1022 & & \Delta^7 0^{10} = 29635200 \\
\Delta^3 0^{10} = 55980 & & \Delta^8 0^{10} = 30240000 \\
\Delta^4 0^{10} = 818520 & & \Delta^9 0^{10} = 16329600 \\
\Delta^5 0^{10} = 5103000 & & \Delta^{10} 0^{10} = 3628800
\end{array}\right. \tag{4.10}$$

$$\left\{\begin{array}{llll}
\pi_{00} = .000362880 & & \pi_{50} = .128595600 \\
\pi_{10} = .016329600 & & \pi_{60} = .017188920 \\
\pi_{20} = .136080000 & & \pi_{70} = .000671760 \\
\pi_{30} = .355622400 & & \pi_{80} = .000004599 \\
\pi_{40} = .345144240 & & \pi_{90} = .000000001
\end{array}\right. \tag{4.11}$$

$$\left\{\begin{array}{llll}
\sigma_1 = 3.486784401 & & m = 3.486784401 \\
\sigma_2 = 9.663676416 & & \sigma^2 = 0.992795358
\end{array}\right. \tag{4.12}$$

---

[7] This result is derived as follows: $(x-a)^r = (1+\Delta)^x\cdot(-a)^r$; $E(x-a)^r = \sum_{x=1}^{n}(x-a)^r f(x) = \left(\sum_{x=1}^{n}(1+\Delta)^x\cdot f(x)\right)\cdot(-a)^r = \left(\sum_{x=1}^{n}(1 + x\Delta + x(x-1)\Delta^2/2! + \cdots)f(x)\right)\cdot(-a)^r$. For a bivariate distribution it may be shown similarly that, symbolically, $E((x-a)^r(y-b)^s) = \{\exp(\sigma_{1.}\,\Delta_1 + \sigma_{.1}\,\Delta_2)\}\cdot(-a)^r(-b)^s$ where $\sigma_{1.}^{\ m}\sigma_{.1}^{\ n} = \sigma_{mn}$ and Δ_1 operates only on a and Δ_2 operates only on b. A similar result may be derived for a multivariate distribution.

[8] cf. Whittaker & Robinson, *op. cit.* p. 7.

The observed distribution was obtained by distributing 200 sets of ten digits each, the digits being found in Tippet's Random Sampling Numbers.[9] The results obtained are given in Fig. 2. Three of the 200 observed sets were illustrated in section 3.

The agreement between observed results and theoretical values is gratifying.

5. **Distribution of the number of events which occur once each.** Let π_{k1}, represent the probability that there are k events which occur once each. Thus, the various probabilities, obtained by rearranging the terms of the expansion of $(p_1 + p_2 + \cdots + p_n)^N$, are as follows:

$$
(5.1)\quad \begin{cases}
\pi_{01} = \sum \dfrac{N!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}, \quad x_1 + x_2 + \cdots + x_n = N, \quad \text{no } x = 1; \\
\pi_{11} = p_1 \sum \dfrac{N!}{x_2! \cdots x_n!} p_2^{x_2} \cdots p_n^{x_n} + \cdots + p_n \sum \dfrac{N!}{x_1! \cdots x_{n-1}!} p_1^{x_1} \cdots p_{n-1}^{x_{n-1}}, \\
\qquad x_1 + x_2 + \cdots + x_{n-1} = N - 1, \text{ etc.}, \quad \text{no } x = 1; \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\pi_{k1} = p_1 p_2 \cdots p_k \sum \dfrac{N!}{x_{k+1}! \cdots x_n!} p_{k+1}^{x_{k+1}} \cdots p_n^{x_n} + \cdots + p_{n-k+1} \cdots p_n \\
\qquad \sum \dfrac{N!}{x_1! \cdots x_{n-k}!} p_1^{x_1} \cdots p_{n-k}^{x_{n-k}}, \\
\qquad x_1 + x_2 + \cdots + x_{n-k} = N - k, \text{ etc.}, \quad \text{no } x = 1;
\end{cases}
$$

No. of events not occurring x	Observed frequency f	Theoretical frequency	xf	$x(x-1)f$	Observed parameters
0	0	0.08	0	0	$\bar{\sigma}_1 = 3.46$
1	8	3.26	8	0	$\bar{\sigma}_2 = 9.61$
2	22	27.22	44	44	$\bar{x} = 3.46$
3	72	71.12	216	432	$s^2 = 1.0984$
4	72	69.02	288	864	Theoretical
5	21	25.72	105	420	Parameters
6	4	3.44	24	120	$\sigma_1 = 3.49$
7	1	0.14	7	42	$\sigma_2 = 9.66$
8	0	0.00	0	0	$m = 3.49$
9	0	0.00	0	0	$\sigma^2 = 0.99$
	200	200.00	692	1922	

FIG. 2

[9] L. H. C. Tippet, Random Sampling Numbers, *Tracts for Computers*, No. XV (1927), London.

In view of (2.21) and (2.27), it is found that (5.1) becomes

$$
(5.2)\quad \begin{cases}
\pi_{01} = 1 - N \sum_{i=1}^{n} p_i(1-p_i)^{N-1} + \dfrac{N(N-1)}{2!} \sum_{i,j=1}^{n} p_i p_j (1 - p_i - p_j)^{N-2} - \cdots \\
\pi_{11} = N\left\{\sum_{i=1}^{n} p_i(1-p_i)^{N-1} - (N-1)\sum_{i,j=1}^{n} p_i p_j(1-p_i-p_j)^{N-2} + \cdots\right\} \\
\pi_{21} = \dfrac{N(N-1)}{2!}\left\{\sum_{i,j=1}^{n} p_i p_j (1-p_i-p_j)^{N-2} - \cdots\right\} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad (i \neq j, \text{ etc.})
\end{cases}
$$

From (5.2) there is readily derived the fact that

$$
(5.3)\quad \sigma_r = N(N-1)\cdots(N-r+1) \sum_{a,b,\cdots,r=1}^{n} p_a p_b \cdots p_r (1 - p_a - p_b - \cdots - p_r)^{N-r}, \quad (a \neq b, \text{ etc.})
$$

For the case in which $p_1 = p_2 = \cdots = p_n = \dfrac{1}{n}$, the distribution in (5.2) becomes

$$
(5.4)\quad \begin{cases}
\pi_{01} = \left(\dfrac{1}{n}\right)^N f_1(n, N) \\
\pi_{11} = \left(\dfrac{1}{n}\right)^N nNf_1(n-1, N-1) \\
\pi_{21} = \left(\dfrac{1}{n}\right)^N \dfrac{n(n-1)N(N-1)}{2!} f_1(n-2, N-2) \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\pi_{r1} = \left(\dfrac{1}{n}\right)^N \dbinom{n}{r} N^{(r)} f_1(n-r, N-r) \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{cases}
$$

where $f_1(n, N)$ and $N^{(r)}$ have been defined in section 2. For this case (5.3) becomes

$$
(5.5)\quad \sigma_r = n^{(r)} N^{(r)} (n-r)^{N-r} / n^N
$$

Evaluation of (5.4) and (5.5) for $n = N = 10$ yields,

$$
(5.6)\quad \begin{cases}
\pi_{01} = .00811639 & \pi_{41} = .27052704 & \pi_{81} = .01632960 \\
\pi_{11} = .04794633 & \pi_{51} = .15621984 & \pi_{91} = .00000000^{10} \\
\pi_{21} = .14082336 & \pi_{61} = .12700800 & \pi_{101} = .00036288 \\
\pi_{31} = .21089376 & \pi_{71} = .02177280 &
\end{cases}
$$

$$
(5.7)\quad \begin{cases}
\sigma_1 = 3.87420489 & m = 3.87420489 \\
\sigma_2 = 13.58954496 & \sigma^2 = 2.45428632
\end{cases}
$$

[10] For the case $n = N = 10$ there cannot be 9 events occurring once each, since then the tenth event must also occur once.

The observed distribution, given in Fig. 3, was obtained from the 200 sets previously considered.

The agreement between the observed results and theoretical values is gratifying.

6. **Distribution of the number of events whlch occur r times each.** Let π_{kr} represent the probability that there are k events occurring r times each. Thus, the various probabilities, obtained by rearranging the terms of the expansion of $(p_1 + p_2 + \cdots + p_n)^N$, are as follows:

No. of events occurring once each x	Observed frequency f	Theoretical frequency	xf	$x(x-1)f$	Observed parameters
0	1	1.62	0	0	$\bar{\sigma}_1 = 3.905$
1	10	9.58	10	0	$\bar{\sigma}_2 = 14.000$
2	30	28.16	60	60	$\bar{x} = 3.905$
3	37	42.18	111	222	$s^2 = 2.656$
4	62	54.10	248	744	Theoretical
5	27	31.24	135	540	Parameters
6	22	25.40	132	660	$\sigma_1 = 3.874$
7	3	4.36	21	126	$\sigma_2 = 13.590$
8	8	3.26	64	448	$m = 3.874$
9	0	0.00	0	0	$\sigma^2 = 2.454$
10	0	0.08	0	0	
	200	199.98	781	2800	

FIG. 3

$$
(6.1)\quad \left\{
\begin{aligned}
\pi_{0r} &= \sum \frac{N!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}, \quad x_1 + x_2 + \cdots + x_n = N, \quad \text{no } x = r; \\
\pi_{1r} &= \frac{p_1^r}{r!} \sum \frac{N!}{x_2! \cdots x_n!} p_2^{x_2} \cdots p_n^{x_n} + \cdots + \frac{p_n^r}{r!} \sum \frac{N!}{x_1! \cdots x_{n-1}!} p_1^{x_1} \cdots p_{n-1}^{x_{n-1}}, \\
&\qquad x_1 + x_2 + \cdots + x_{n-1} = N - r, \text{ etc.}, \quad \text{no } x = r; \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\pi_{kr} &= \frac{p_1^r p_2^r \cdots p_k^r}{(r!)^k} \sum \frac{N!}{x_{k+1}! \cdots x_n!} p_{k+1}^{x_{k+1}} \cdots p_n^{x_n} + \cdots \\
&\qquad + \frac{p_{n-k+1}^r \cdots p_n^r}{(r!)^k} \sum \frac{N!}{x_1! \cdots x_{n-k}!} p_1^{x_1} \cdots p_{n-k}^{x_{n-k}}, \\
&\qquad x_1 + x_2 + \cdots + x_{n-k} = N - kr, \text{ etc.}, \quad \text{no } x = r; \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
\right.
$$

In view of (2.21) and (2.27) it is found that (6.1) becomes

$$(6.2)\quad \begin{cases} \pi_{0r} = 1 - \dfrac{N^{(r)}}{r!}\displaystyle\sum_{i=1}^{n} p_i^r(1-p_i)^{N-r} + \dfrac{N^{(2r)}}{2!\,(r!)^2}\displaystyle\sum_{i,j=1}^{n} p_i^r p_j^r(1-p_i-p_j)^{N-2r} - \cdots \\ \pi_{1r} = \dfrac{N^{(r)}}{r!}\left\{\displaystyle\sum_{i=1}^{n} p_i^r(1-p_i)^{N-r} - \dfrac{(N-r)^{(r)}}{r!}\displaystyle\sum_{i,j=1}^{n} p_i^r p_j^r(1-p_i-p_j)^{N-2r} + \cdots\right\} \\ \pi_{2r} = \dfrac{N^{(2r)}}{2!\,(r!)^2}\left\{\displaystyle\sum_{i,j=1}^{n} p_i^r p_j^r(1-p_i-p_j)^{N-2r} - \cdots\right\} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad\qquad (i \neq j, \text{etc.}) \end{cases}$$

From (6.2) there is readily derived the fact that

$$(6.3)\quad \sigma_k = \frac{N^{(kr)}}{(r!)^k}\sum_{a,b,\cdots,k=1}^{n} p_a^r p_b^r \cdots p_k^r(1-p_a-p_b-\cdots-p_k)^{N-kr}, \quad (a \neq b, \text{etc.})$$

For $r = 0,1$ (6.2) and (6.3) reduce to the values previously derived.

For the case in which $p_1 = p_2 = \cdots = p_n = \dfrac{1}{n}$, the distribution in (6.2) becomes

$$(6.4)\quad \begin{cases} \pi_{0r} = \left(\dfrac{1}{n}\right)^N f_r(n, N) \\ \pi_{1r} = \left(\dfrac{1}{n}\right)^N \dfrac{nN^{(r)}}{r!} f_r(n-1, N-r) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \pi_{kr} = \left(\dfrac{1}{n}\right)^N \dbinom{n}{k} \dfrac{N^{(kr)}}{(r!)^k} f_r(n-k, N-kr) \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

where $f_r(n, N)$ has been defined in section 2. For this case (6.3) becomes

$$(6.5)\quad \sigma_k = N^{(kr)} n^{(k)}(n-k)^{N-kr}/n^N$$

7. Simultaneous distribution of the number of events not occurring, and of the number of events occurring once each. The probabilities for the simultaneous occurrence of the various combinations of the number of events not occurring, and of the number of events occurring once each, are given by rearranging the terms of the expansion of $(p_1 + p_2 + \cdots + p_n)^N$, and are given as in Fig. 4.

In Fig. 4 none of the subscripts take on equal values simultaneously, and G_{01} has been defined in section 2. Summation of the values in the k-th column of Fig. 4, yields the probability that there are $(k - 1)$ events not occurring. Comparison with (4.2) yields

$$(7.1)\quad \begin{aligned} F_0(n, N, p_1, p_2, \cdots, p_n) &= G_0(n, N) = G_{01}(n, N) + N\sum_{i=1}^{n} p_i G_{01}(n-1, N-1, p_i) \\ &+ \frac{N^{(2)}}{2!}\sum_{i,j=1}^{n} p_i p_j G_{01}(n-2, N-2, p_i, p_j) + \cdots, \quad (i \neq j, \text{etc.}) \end{aligned}$$

Number of events occurring once each		Number of events not occurring: 0	1	r
	0	$G_{01}(n, N)$	$\sum_{i=1}^{n} G_{01}(n-1, N, p_i)$	$\cdots$
	1	$N\sum_{i=1}^{n} p_i G_{01}(n-1, N-1, p_i)$	$N\sum_{i,j=1}^{n} p_i G_{01}(n-2, N-1, p_i, p_j)$	$\cdots$
	2	$\frac{N^{(2)}}{2!}\sum_{i,j=1}^{n} p_i p_j G_{01}(n-2, N-2, p_i p_j)$	$\frac{N^{(2)}}{2!}\sum_{i,j,k=1}^{n} p_i p_j G_{01}(n-3, N-2, p_i, p_j, p_k)$	$\cdots$
	s	$\cdots$	$\cdots$	$\frac{N^{(s)}}{r!\,s!}\sum_{a,b,\cdots,s,\alpha,\beta,\cdots,\rho=1}^{n} p_a p_b \cdots p_s G_{01}(n-r-s, N-s, p_a, \cdots, p_s, p_\alpha, \cdots, p_\rho)$

FIG. 4

Summation of the values in the k-th row of Fig. 4, yields the probability that there are $(k - 1)$ events occurring once each. Comparison with (5.2) and (2.27) yields

$$F_1(n, N, p_1, p_2, \cdots, p_n) = G_1(n, N) = G_{01}(n, N) + \sum_{i=1}^{n} G_{01}(n-1, N, p_i) + \frac{1}{2!}\sum_{i,j=1}^{n} G_{01}(n-2, N, p_i, p_j) + \cdots, \quad (i \neq j, \text{etc.}) \tag{7.2}$$

If we use x to represent the number of events not occurring, and y the number of events occurring once each, then it is found that

$$E(x^{(r)} y^{(s)}) = \sigma_{rs} = N^{(s)} \sum_{a,b,\cdots,s,\alpha,\beta,\cdots,\rho=1}^{n} p_a p_b \cdots p_s(1 - p_a - \cdots - p_s - p_\alpha - \cdots - p_\rho)^{N-s}, \quad (a \neq b, \text{etc.}). \tag{7.3}$$

If ${}_0\bar{x}_{k1}$ represents the average number of events not occurring, when there are k events occurring once each, then from Fig. 4 there is found that

$$ {}_0\bar{x}_{01} = \frac{\sum_{i=1}^{n} G_{01}(n-1, N, p_i) + 2\sum_{i,j=1}^{n} G_{01}(n-2, N, p_i, p_j)/2! + 3\sum_{i,j,k=1}^{n} G_{01}(n-3, N, p_i, p_j, p_k)/3! + \cdots}{G_{01}(n, N) + \sum_{i=1}^{n} G_{01}(n-1, N, p_i) + \sum_{i,j=1}^{n} G_{01}(n-2, N, p_i, p_j)/2! + \cdots} \quad (i \neq j, \text{etc.}) \tag{7.4}$$

In view of (7.2), (7.4) reduces to

$$ {}_0\bar{x}_{01} = \left(\sum_{i=1}^{n} G_1(n, N, p_i)\right) \Big/ G_1(n, N) \tag{7.5}$$

A similar procedure, yields, in general

$$ {}_0\bar{x}_{k1} = \frac{\displaystyle\sum_{a,b,\cdots,k,l=1}^{n} p_a p_b \cdots p_k G_1(n-k-1, N-k, p_a, p_b, \cdots, p_k, p_l)}{\displaystyle\sum_{a,b,\cdots,k=1}^{n} p_a p_b \cdots p_k G_1(n-k, N-k, p_a, p_b, \cdots, p_k)} \tag{7.6}$$

$(a \neq b, \text{etc.})$

If ${}_1\bar{y}_{k0}$ represents the average number of events occurring once each, when there are k events not occurring, then from Fig. 4, there is found that

$$ {}_1\bar{y}_{00} = \frac{N\left\{\displaystyle\sum_{i=1}^{n} p_i G_{01}(n-1, N-1, p_i) + 2(N-1) \sum_{i,j=1}^{n} p_i p_j G_{01}(n-2, N-2, p_i, p_j)/2! + \cdots\right\}}{G_{01}(n, N) + N \displaystyle\sum_{i=1}^{n} p_i G_{01}(n-1, N-1, p_i) + N^{(2)} \sum_{i,j=1}^{n} p_i p_j G_{01}(n-2, N-2, p_i, p_j)/2!} \quad (i \neq j, \text{etc.}) \tag{7.7}$$

In view of (7.1), (7.7) reduces to

$$ {}_1\bar{y}_{00} = \left(N\sum_{i=1}^{n} p_i G_0(n-1, N-1, p_i)\right) \Big/ G_0(n, N) \tag{7.8}$$

A similar procedure, yields, in general

$$ {}_1\bar{y}_{k0} = \frac{N \displaystyle\sum_{a,b,\cdots,k,l=1}^{n} p_a G_0(n-k-1, N-1, p_a, p_b, \cdots, p_k, p_l)}{\displaystyle\sum_{a,b,\cdots,k=1}^{n} G_0(n-k, N, p_a, p_b, \cdots, p_k)} \quad (a \neq b, \text{etc.}) \tag{7.9}$$

For the case in which $p_1 = p_2 = \cdots = p_n = \dfrac{1}{n}$, as may be found from Fig. 4, the probability for the simultaneous occurrence of r events not occurring, and s events occurring once each, is given by

$$ \left(\frac{1}{n}\right)^N \frac{n^{(r+s)} N^{(s)}}{r!\, s!} f_{01}(n-r-s, N-s) \tag{7.10}$$

For this case (7.1), (7.2), (7.3), (7.6), and (7.9) yield respectively

$$ f_0(n, N) = f_{01}(n, N) + nNf_{01}(n-1, N-1) + \binom{n}{2} N^{(2)} f_{01}(n-2, N-2) + \cdots \tag{7.11}$$

$$ f_1(n, N) = f_{01}(n, N) + nf_{01}(n-1, N) + \binom{n}{2} f_{01}(n-2, N) + \cdots \tag{7.12}$$

$$ \sigma_{rs} = N^{(s)} n^{(r+s)} (n-r-s)^{N-s}/n^N \tag{7.13}$$

$$(7.14)\qquad {}_0\bar{x}_{k1} = (n - k)f_1(n - k - 1, N - k)/f_1(n - k, N - k)$$

$$(7.15)\qquad {}_1\bar{y}_{k0} = N(n - k)f_0(n - k - 1, N - 1)/f_0(n - k, N)$$

Let us consider again the case when $p_1 = p_2 = \cdots = p_n = \dfrac{1}{n}$ and $n = N = 10$. Evaluating (7.14) and (7.15) by means of (2.15) yields

$$(7.16)\qquad \begin{cases} {}_0\bar{x}_{01} = 5.71 & {}_0\bar{x}_{51} = 3.02 \\ {}_0\bar{x}_{11} = 5.21 & {}_0\bar{x}_{61} = 2.10 \\ {}_0\bar{x}_{21} = 4.51 & {}_0\bar{x}_{71} = 2.00 \\ {}_0\bar{x}_{31} = 4.10 & {}_0\bar{x}_{81} = 1.00 \\ {}_0\bar{x}_{41} = 3.28 & {}_0\bar{x}_{91} = 0.00 \end{cases}$$

$$(7.17)\qquad \begin{cases} {}_1\bar{y}_{00} = 10.00 & {}_1\bar{y}_{50} = 1.83 \\ {}_1\bar{y}_{10} = 8.00 & {}_1\bar{y}_{60} = 0.89 \\ {}_1\bar{y}_{20} = 6.16 & {}_1\bar{y}_{70} = 0.27 \\ {}_1\bar{y}_{30} = 4.50 & {}_1\bar{y}_{80} = 0.02 \\ {}_1\bar{y}_{00} = 3.05 & {}_1\bar{y}_{90} = 0.00 \end{cases}$$

The 200 sets of observations already considered yielded the simultaneous distribution given in Fig. 5.

Number of events occurring once each		Number of events not occurring											
		0	1	2	3	4	5	6	7	8	9		$\bar{x}$
	0								1			1	7.00
	1					1	6	3				10	5.20
	2					16	13	1				30	4.50
	3					35	2					37	4.05
	4				42	20						62	3.32
	5				27							27	3.00
	6			19	3							22	2.14
	7			3								3	2.00
	8		8									8	1.00
	9											0	
	10											0	
		0	8	22	72	72	21	4	1	0	0	200	
	$\bar{y}$		8.00	6.16	4.46	3.03	1.81	1.25	0.00				

Fig. 5

The distribution in Fig. 5 yields $\bar{\sigma}_{11} = 11.89$, (7.13) yields $\sigma_{11} = 12.07959552$.

The agreement between the observed results in Fig. 5 and the theoretical values in (7.16) and (7.17) is gratifying.

8. **Simultaneous distribution of the number of events which occur r times each, and of the number of events which occur s times each.** The probabilities for the simultaneous occurrence of the various combinations of the number of events which occur r times each, and of the number of events which occur s times each, are obtained by rearranging the terms of the expansion of $(p_1 + p_2 + \cdots + p_n)^N$. If $\pi_{kr,ls}$ is the probability for the simultaneous occurrence of k events which occur r times each and l events which occur s times each, then

$$\begin{aligned}\pi_{kr,ls} = \frac{N^{(kr+ls)}}{k!\,l!\,(r!)^k\,(s!)^l} \sum_{a,b,\cdots,k,\alpha,\beta,\cdots,\lambda=1}^{n} p_a^r \cdots p_k^r p_\alpha^s \cdots p_\lambda^s G_{rs} \qquad (8.1)\\ (n-k-l, N-kr-ls, p_a, \cdots, p_k, p_\alpha, \cdots, p_\lambda), \qquad (a \neq b, \text{ etc.})\end{aligned}$$

where G_{rs} is defined in section 2.

From (8.1) and (6.2), there is derived, in a manner similar to the derivation of (7.1) and (7.2), the result that

$$\begin{aligned}F_r(n, N, p_1, \cdots, p_n) = G_r(n, N) = G_{rs}(n, N) + \frac{N^{(s)}}{s!} \sum_{i=1}^{n} p_i^s G_{rs}(n-1, N-s, p_i) \qquad (8.2)\\ + \frac{N^{(2s)}}{2!\,(s!)^2} \sum_{i,j=1}^{n} p_i^s p_j^s G_{rs}(n-2, N-2s, p_i, p_j) + \cdots, \qquad (i \neq j, \text{ etc.})\end{aligned}$$

and a similar result by interchanging r and s in (8.2).

For the distribution given by (8.1), it is found that

$$\begin{aligned}\sigma_{kl} = \frac{N^{(kr+ls)}}{(r!)^k\,(s!)^l} \sum_{a,b,\cdots,k,\alpha,\beta,\cdots,\lambda=1}^{n} p_a^r \cdots p_k^r p_\alpha^s \cdots p_\lambda^s \qquad (8.3)\\ (1 - p_a - \cdots - p_k - p_\alpha - \cdots - p_\lambda)^{N-kr-ls}, \qquad (a \neq b, \text{ etc.})\end{aligned}$$

If ${}_r\bar{x}_{ls}$ represents the average number of events which occur r times each when there are l events which occur s times each, then from (8.1) and (8.2), in a manner similar to the derivation of (7.6), it is found that

$$ {}_r\bar{x}_{ls} = \frac{(N-ls)^{(r)} \displaystyle\sum_{a,\alpha,\cdots,\lambda=1}^{n} p_a^r p_\alpha^s \cdots p_\lambda^s G_s(n-1-l, N-r-ls, p_a, p_\alpha, \cdots, p_\lambda)}{r! \displaystyle\sum_{\alpha,\cdots,\lambda=1}^{n} p_\alpha^s \cdots p_\lambda^s G_s(n-l, N-ls, p_\alpha, \cdots, p_\lambda)} \qquad (8.4)$$

$$(\alpha \neq \beta, \text{ etc.})$$

If ${}_s\bar{y}_{kr}$ represents the average number of events which occur s times each when there are k events which occur r times each, then by interchanging k and l, and r and s in (8.4), there is found

$$
{}_s\bar{y}_{kr} = \frac{(N-kr)^{(s)} \displaystyle\sum_{a,\cdots,k,\alpha=1}^{n} p_a^r \cdots p_k^r p_\alpha^s G_r(n-k-1, N-kr-s, p_a, \cdots, p_k, p_\alpha)}{\displaystyle\sum_{a,b,\cdots,k=1}^{n} p_a^r \cdots p_k^r G_r(n-k, N-kr, p_a, \cdots, p_k)} \tag{8.5}
$$

$(a \neq b, \text{etc.})$

For the case when $p_1 = p_2 = \cdots = p_n = \dfrac{1}{n}$, it is found that (8.1), (8.2), (8.3), (8.4), and (8.5) respectively yield

$$
\pi_{kr,ls} = \left(\frac{1}{n}\right)^N \frac{n^{(k+l)} N^{(kr+ls)}}{k!\,l!\,(r!)^k\,(s!)^l} f_{rs}(n-k-l, N-kr-ls) \tag{8.6}
$$

$$
\begin{aligned}
f_r(n, N) = f_{rs}(n, N) + \frac{nN^{(s)}}{s!} f_{rs}(n-1, N-s) \\
+ \frac{n(n-1)N^{(2s)}}{2!\,(s!)^2} f_{rs}(n-2, N-2s) + \cdots
\end{aligned} \tag{8.7}
$$

$$
\sigma_{kl} = n^{(k+l)} N^{(kr+ls)} (n-k-l)^{N-kr-ls} / (r!)^k\,(s!)^l\,n^N \tag{8.8}
$$

$$
{}_r\bar{x}_{ls} = (n-l)(N-ls)^{(r)} f_s(n-1-l, N-r-ls)/r!\,f_s(n-l, N-ls) \tag{8.9}
$$

$$
{}_s\bar{y}_{kr} = (n-k)(N-kr)^{(s)} f_r(n-k-1, N-kr-s)/s!\,f_r(n-k, N-kr) \tag{8.10}
$$

For $r = 0$, $s = 1$, the results derived in this section of course reduce to those already derived in section 7.

9. **Conclusion.** It is clear that the same method of procedure may be employed to study the simultaneous distribution of the number of events which occur r, s, $\cdots$, t, times each. However we will not continue the discussion any further.

We have thus seen that the multinomial distribution serves as the background for the study of a number of distributions which have certain practical applications.

The theory discussed herein has been illustrated by several examples which yielded gratifying agreement between observed and theoretical results.

WASHINGTON, D. C.

A PROBLEM IN LEAST SQUARES

By Jan K. Wiśniewski

§1. We are dealing with two variables, the observed values of which are denoted x and y respectively. The pairs of observations are divided into r groups, numbering $n_1, n_2, \cdots n_r$ pairs. Suppose in each group we determine a regression equation of the following shape:

$$y_i = a_i + b_i x + \cdots m_i x^s \tag{1}$$

where y_i denotes the value of the "dependent" variable obtained from the regression equation, while y without any subscript denotes its observed value. The r regression equations of type (1) are not assumed independent; on the contrary, we postulate that

$$\sum_1^r y_i = a_0 + b_0 x + \cdots m_0 x^s \tag{2}$$

be fulfilled identically in x; $a_0, b_0, \cdots m_0$ being predetermined numbers. This leads to the following conditions:

$$\sum_1^r a_i = a_0 \qquad \sum_1^r b_i = b_0 \cdots \sum_1^r m_i = m_0. \tag{3}$$

The magnitude to be minimized under the theory of least squares is now

$$\begin{aligned} Z = \sum_1^{r-1} \textstyle\sum_i [y - (a_i + b_i x + \cdots m_i x^s)]^2 + \textstyle\sum_r \Big\{ y - \Big[\Big(a_0 - \displaystyle\sum_1^{r-1} a_i \Big) \\ + \Big(b_0 - \sum_1^{r-1} b_i \Big) x + \cdots \Big(m_0 - \sum_1^{r-1} m_i \Big) x^s \Big] \Big\}^2. \end{aligned} \tag{4}$$

The normal equations derived from (4) are of the following shape:

. . . .

$$\begin{aligned} n_j a_j + n_r \sum_1^{r-1} a_i + b_j \textstyle\sum_j x + \Big(\displaystyle\sum_1^{r-1} b_i \Big) (\textstyle\sum_r x) + \cdots m_j \textstyle\sum_j x^s \\ + \Big(\sum_1^{r-1} m_i \Big) (\textstyle\sum_r x^s) = \textstyle\sum_j y - \textstyle\sum_r y + n_r a_0 + b_0 \textstyle\sum_r x + \cdots m_0 \textstyle\sum_r x^s \end{aligned} \tag{5}$$

. . . .

. . . .

$$a_j \sum_j x + \left(\sum_1^{r-1} a_i\right) (\sum_r x) + b_j \sum_j x^2 + \left(\sum_1^{r-1} b_i\right) (\sum_r x^2)$$
$$+ \cdots m_j \sum_j x^{s+1} + \left(\sum_1^{r-1} m_i\right) (\sum_r x^{s+1}) = \sum_j xy - \sum_r xy + a_0 \sum_r x$$
$$+ b_0 \sum_r x^2 + \cdots m_0 \sum_r x^{s+1}$$

....

.... (5)

....

$$a_j \sum_j x^s + \left(\sum_1^{r} a_i\right) (\sum_r x^s) + b_j \sum_j x^{s+1} + \left(\sum_1^{r-1} b_i\right) (\sum_r x^{s+1})$$
$$+ \cdots m_j \sum_j x^{2s} + \left(\sum_1^{r-1} m_i\right) (\sum_r x^{2s}) = \sum_j x^s y - \sum_r x^s y$$
$$+ a_0 \sum_r x^s + b_0 \sum_r x^{s+1} + \cdots m_0 \sum_r x^{2s}$$

....

$\sum_i$ meaning a summation extended over the i-th group. As (1) is of the s-th degree, we have $(s + 1)$ $(r - 1)$ parameters to determine and as many equations, the problem thus being in theory solved.* As to the numerical solution, Doolittle's method or any other may be applied. We do not enter at present the question, how much labor would the actual solution require.

Examples. Allen and Bowley in their book on "Family Expenditure" (London, 1935) assume the expenditure on some defined item f to be a linear function of the total expenditure e

$$f = ke + c. \tag{6}$$

Evidently $\sum k = 1$, $\sum c = 0$ (cfr. pp. 10–11). Another example I give in a paper on seasonal variation, which appeared in "Economic Studies" III (Kraków). Actual values y of a time series are assumed to be linear functions of certain "normal" values x

$$y = a + bx \tag{7}$$

a and b changing from month to month but constant from year to year. Then $\sum a = 0$, $\sum b = 12$.

§2. **Methods of solution in special cases.** The generally recognized methods of solving normal equations become extremely laborious as the product $(s + 1)$ $(r - 1)$ grows large. As a matter of fact, the amount of computer's work is approximately proportional to the cube of the number of parameters to determine. Therefore short cuts seem to be indispensable. A most elegant one is at our disposal in the special case[1] when the values of x in the several groups

* The remaining $s + 1$ parameters a_r, b_r, $\cdots$ m are, of course, found from (3).

[1] This seems to be realized in Allen and Bowley's work.

are identical, or, at least, the sums n_i, $\sum_i x$, $\sum_i x^2$, $\cdots \sum_i x^{2s}$ are identical in i. Instead of (1) we shall write

$$y_i = A_i + B_i X_1 + \cdots M X_s \tag{8}$$

where $X_1, X_2, \cdots X_s$ are orthogonal polynomials, i.e. such that $\sum X_i X_j = 0$ if and only if $i \neq j$. In general, $X_h = X^h + \alpha_{h-1}^h X^{h-1} + \cdots \alpha_0^h$, the coefficients being rational functions of n, $\sum x$, $\sum x^2$, $\cdots \sum x^{2s-1}$.

The conditions (3) can now be replaced by a set of equivalent ones, viz.

$$\sum_1^r A_i = A_0 \qquad \sum_1^r B_i = B_0 \cdots \sum_1^r M_i = M_0. \tag{9}$$

How the actual values of A_0, B_0, $\cdots M_0$ are found, will be shown in the next paragraph. The solution becomes now very easy, as the normal equations for the determination of each set of $r - 1$ parameters are independent, i.e. we can calculate the A's separately, then the B's etc., the order of solution being of no importance. Moreover the shape of the normal equations permits of considerable simplification of solution. Suppose we have to determine the values of the coefficients K, corresponding to X_h. The normal equations are now—after certain simplifications—

$$\begin{aligned}
2K_1 + K_2 + K_3 + \cdots K_{r-1} &= \frac{1}{\sum X_h^2}\left(\textstyle\sum_1 X_h y - \sum_r X_h y\right) + K_0 \\
K_1 + 2K_2 + K_3 + \cdots K_{r-1} &= \frac{1}{\sum X_h^2}\left(\textstyle\sum_2 X_h y - \sum_r X_h y\right) + K_0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
K_1 + K_2 + K_3 + \cdots 2K_{r-1} &= \frac{1}{\sum X_h^2}\left(\textstyle\sum_{r-1} X_h y - \sum_r X_h y\right) + K_0.
\end{aligned} \tag{10}$$

Adding these equations, dividing the sum by r and substracting the quotient from the j-th equation, we get

$$K_j = \frac{\sum_j X_h y}{\sum X_h^2} - \frac{1}{r}\left(\sum_1^r \frac{\sum_i X_h y}{\sum X_h^2} - K_0\right). \tag{11}$$

The first member of the right hand side of (11) should be regarded as the principal term: this is actually the value we would obtain for K_j, were this coefficient independent from the other K's. The second member is a correction term, the necessary amount of correction being distributed equally among the several K's. The simple solution given by (11) is only possible if the sum $\sum X_h^2$ is the same for each group. From the definition of X_h we see that it is equivalent to saying that n_i, $\sum_i x$, $\sum_i x^2$, $\cdots \sum_i x^{2h}$ be identical in i. As h increases to s, we come to the condition given at the beginning of this paragraph.

§3. If this condition is not fulfilled, we can, indeed, replace the power series in x by orthogonal polynomials $X_{h \cdot i}$, the second subscript being appended in order to show that the values of the X polynomials are no more identical for the several groups; these polynomials are now orthogonalized separately within each group. But we are no more able to predetermine the values of $A_0, B_0, \cdots M_0$, as they depend on each other; this will be made clear a little later. Therefore we have to resort to an approximation: the values of the parameters will not be found from simultaneous equations, but successively, step by step, beginning with those corresponding to the highest degree of the independent variable.

The values of $a_0, b_0, \cdots m_0$ are given. It is evident that $m_0 = M_0$. The j-th normal equation is now:

$$M_j \sum_j X_{s \cdot j}^2 - M_0 \sum_r X_{s \cdot r}^2 + \left(\sum_1^{r-1} M_i\right)\left(\sum_r X_{s \cdot r}^2\right) = \sum_j X_{s \cdot j} y - \sum_r X_{s \cdot j} y. \tag{12}$$

We see at once that

$$M_i = \frac{M_j \sum_j X_{s \cdot j}^2 + \sum_i X_{s \cdot i} y - \sum_j X_{s \cdot j} y}{\sum_i X_{s \cdot i}^2}. \tag{13}$$

Inserting this into /12/ we get

$$M_j = \frac{\sum_j X_{s \cdot j} y}{\sum_j X_{s \cdot j}^2} - \frac{1}{\sum_j X_{s \cdot j}^2} \cdot \frac{\sum_1^r \frac{\sum_i X_{s \cdot i} y}{\sum_i X_{s \cdot i}^2} - M_0}{\sum_1^r \frac{1}{\sum_i X_{s \cdot i}^2}}. \tag{14}$$

The second member of the right hand side of /14/ is again a correction term, the necessary amount of correction being distributed in inverse proportion to $\sum_j X_{s \cdot j}^2$. Now we determine the value of L_0, this coefficient corresponding to $s - 1$, the second highest degree of x, and calculate the several L's from equations strictly analogous to (14) thus accomplishing the second step of our work, and so on, down to the A's. L_0 is found from the following equation:

$$L_0 = l_0 - \sum_1^r [\alpha_{s-1}^s(i) \cdot M_i]. \tag{15}$$

To α_{s-1}^s is now appended a bracketed i, this to stress its variation from group to group. We see from (15) that before the several M's are calculated we are not in a position to determine L_0. On the other hand, if α_{s-1}^s is the same for all groups, the second member of the right hand side of (15) simply reduces to $\alpha_{s-1}^s \cdot m_0$ and L_0 can be determined in advance, i.e. before calculating the M's. This is the case treated first (in §2). In any case, if no definite correlation is to be expected between $\alpha_{s-1}^s(i)$ and M_i, the approximative method developed here should give very nearly correct results. The writer applied this method of solution to the simple problem of seasonal variation mentioned in §1 and found the results very satisfactory.

A SIGNIFICANCE TEST FOR COMPONENT ANALYSIS

By Paul G. Hoel

1. Introduction

During the last few years several papers and books have been written on various aspects of what has been termed component or factor analysis. This analysis has arisen from the psychological problem of describing the results on a series of tests in terms of a few distinct abilities or components. In much of such work it is claimed that there does not exist more than a certain number of components, the material discarded in order to substantiate such a claim being considered as due to random errors of sampling or errors of measurement. However, mere inspection of results or the calculation of standard errors of residual correlations is hardly sufficient to justify such conclusions, and therefore a significance test of some kind is necessary. Hotelling[1] considered such a test but based it upon an uncertain analogy with the analysis of variance and upon the legitimacy of using standard errors. The purpose of this paper is to derive a test which is more general in scope and in which all assumptions are explicitly stated.

If each test score is thought of as being made up of two parts, a true score and an error element, the assumption that there exists fewer components than the number of tests implies that the scatter diagram of the true scores will lie in a space of correspondingly smaller dimensionality. Consequently, an ideal test for the number of components would be one which would test the rank of the true moment matrix. In the case of normally distributed variables, this line of approach leads one to the sampling distribution of the generalized variance. Unfortunately, this distribution appears in unintegrated form; however, by considering its moments it is possible to find a good approximation to this exact distribution for samples which are not too small.

The paper proceeds by first finding two approximation distributions for the generalized variance, one for samples which are not too small and one for large samples. It then considers the type of population from which it will be assumed the sample was drawn, and finally applies the test to two numerical examples from recent literature along such lines.

2. Approximation Distributions

Suppose that N individuals have been drawn at random from an n variate normal population whose distribution is expressed by

$$P(x_1, x_2, \cdots, x_n) = Ke^{-\sum_1^n A_{ij}x_ix_j} \tag{1}$$

[1] Harold Hotelling, Analysis of a Complex of Statistical Variables into Principal Components, The Journal of Educational Psychology, September and October, 1933, pp. 21–25.

where $x_i = X_i - m_i$, $A_{ij} = \frac{\Delta_{ij}}{2\sigma_i\sigma_j\Delta}$, Δ is the determinant $|\rho_{ij}|$ and Δ_{ij} is the cofactor of ρ_{ij} in Δ, and $K = |A_{ij}|^{\frac{1}{2}}/(2\pi)^{n/2}$. If the observed values of the variables of the αth individual are denoted by $X_{i\alpha}(i = 1, 2, \cdots, n)$, then the generalized sample variance is defined as $z = |a_{ij}|$, where $a_{ij} = \frac{1}{N}\sum_{\alpha=1}^{N}(X_{i\alpha} - \bar{X}_i)(X_{j\alpha} - \bar{X}_j)$. Wilks[2] has shown that in sampling from the population (1), the kth moment of the sampling distribution of z is given by

$$M_k = A^{-k}\frac{\Gamma\left(\frac{N+2k-1}{2}\right)\Gamma\left(\frac{N+2k-2}{2}\right)\cdots\Gamma\left(\frac{N+2k-n}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)\cdots\Gamma\left(\frac{N-n}{2}\right)}$$

where $A = N^n|A_{ij}|$. An inspection of the integrated form of the distribution of z in the case of $n = 1$ and $n = 2$ suggests that there likely exists a function of similar form for higher values of n whose kth moment can be made to differ from M_k only in higher powers of terms which contain N^{-1} as a factor. An investigation along such lines leads to the function

$$g(z) = Cz^m e^{-n\sqrt[n]{az}} \tag{2}$$

where $C = \frac{a^{\frac{N-n}{2}} n^{n\frac{N-n}{2}-1}}{\Gamma\left(n\frac{N-n}{2}\right)}$, $m = \frac{N-n-2}{2}$, $a = Aq$ and $q = 1 - \frac{(n-1)(n-2)}{2N}$.

It will be shown that the kth moment M'_k of $g(z)$ differs from M_k only in terms of magnitude less than the second and higher powers of k^2n/N or kn^2/N.

Multiplying $g(z)$ by z^k and integrating over the entire range of z will yield M'_k, which turns out to be

$$M'_k = \frac{\Gamma\left(n\frac{N-n+2k}{2}\right)}{a^k n^{nk}\Gamma\left(n\frac{N-n}{2}\right)}.$$

Upon reducing the upper gamma function and performing successive steps of simple algebra

$$\begin{aligned}M'_k &= a^{-k}n^{-nk}\left(n\frac{N-n+2k}{2}-1\right)\left(n\frac{N-n+2k}{2}-2\right)\cdots\left(n\frac{N-n}{2}\right)\\ &= N^{nk}a^{-k}2^{-nk}\left(1+\frac{2k-n-2/n}{N}\right)\left(1+\frac{2k-n-4/n}{N}\right)\cdots\\ &\qquad\left(1+\frac{2k-n-2kn/n}{N}\right).\end{aligned}$$

[2] S. S. Wilks, Certain Generalizations in the Analysis of Variance, Biometrika, Vol. XXIV, 1923, p. 477.

The terms in parentheses may be treated as the factored form of a polynomial of the nkth degree in unity. Thus the quantities $\frac{2k - n - 2/n}{N}$, etc., may be treated as the zeros with signs changed of the corresponding polynomial in x (say). As a result, the successive terms after the first in the non-factored form of this polynomial in unity are the sums of the products of these quantities taken one at a time, two at a time, etc. Upon performing this multiplication and letting $\phi = N^n/2^n A$, M'_k assumes the form

$$M'_k = \phi^k q^{-k}\left[1 - \frac{k(n^2 - nk + 1)}{N} + \cdots\right]$$

where the neglected terms are in magnitude less than the second and higher powers of k^2n/N or kn^2/N. If M_k is handled in exactly the same manner, it will be found that

$$\begin{aligned} M_k = A^{-k}&\left(\frac{N + 2k - 1}{2} - 1\right)\cdots\left(\frac{N + 2k - 1}{2} - k\right)\cdots \\ &\qquad\left(\frac{N + 2k - n}{2} - 1\right)\cdots\left(\frac{N + 2k - n}{2} - k\right) \\ = N^{nk}A^{-k}2^{-nk}&\left(1 + \frac{2k - 3}{N}\right)\cdots\left(1 - \frac{1}{N}\right)\cdots \\ &\qquad\left(1 + \frac{2k - n - 2}{N}\right)\cdots\left(1 - \frac{n}{N}\right) \\ = \phi^k&\left[1 - \frac{nk(n - 2k + 3)}{2N} + \cdots\right] \end{aligned}$$

where the neglected terms are of the same order of magnitude as those neglected in the approximation to M'_k. Before a comparison of M_k and M'_k is possible, the factor q^{-k} of M'_k must be expanded and multiplied into the quantity in brackets. This operation yields the result

$$M'_k = \phi^k\left[1 - \frac{nk(n - 2k + 3)}{2N} + \cdots\right].$$

Thus M_k and M'_k agree to within neglected terms. As a matter of fact, if the values of the neglected terms are considered more carefully, it will be found that the actual difference between M_k and M'_k is considerably less than the given upper bound for the magnitude of neglected terms would indicate. For example, when $n = 5$ the first term in the difference is $6k(k - .9)N^{-2}$, while $625k^2N^{-2}$ or $25k^4N^{-2}$ is the upper bound for this term when only general results are used. The general formula for the first term in this difference has been obtained, but since the remaining terms have not been investigated and since the type of problems to which the distribution $g(z)$ is to be applied does not

justify this refinement, it will not be considered here. Consequently, if one considers this distribution function as sufficiently determined by its low order moments and if one applies $g(z)$ only to problems in which N is fairly large compared with n^2, then the function $g(z)$ will give a good approximation to the exact sampling distribution of z. Obviously, $g(z)$ is identical with the exact distribution for the known cases of $n = 1$ and $n = 2$. It is not possible under the above expansions to vary the constants in the form of $g(z)$ in such a manner as to obtain an approximation whose kth moment will agree with M_k to within still higher powers of comparable terms.

In order to test whether or not a sample value $z = Z$ can be reasonably assumed to have been obtained in random sampling from a population of type (1) with fixed A, it is necessary to calculate the probability P of obtaining in repeated samples a value of z greater than Z. Thus it is necessary to evalua te

$$P = 1 - \int_0^z g(z)\,dz.$$

Upon making the substitution $x = n\sqrt[n]{az}$, and letting $p = n\dfrac{N-n}{2} - 1$ and $u = n\sqrt[n]{aZ}\left(n\dfrac{N-n}{2}\right)^{-\frac{1}{2}} = nN\sqrt[n]{\dfrac{Z}{\phi}\left[1 - \dfrac{(n-1)(n-2)}{2N}\right]}\,[2n(N-n)]^{-\frac{1}{2}}$, this integral can be reduced to the standard form of the incomplete gamma function. Hence P assumes the form

$$P = 1 - I(u, p) \tag{3}$$

where

$$I(u, P) = \frac{1}{\Gamma(p+1)} \int_0^{u\sqrt{p+1}} e^{-x} x^p\,dx.$$

In many applications of this distribution it will be found that the values of u and p lie beyond the tabled[3] values of these constants. Consequently, it will often be sufficient to use the normal distribution to which the gamma distribution tends as N becomes large. This normal distribution will be considered next.

Rather than obtain a normal approximation to $g(z)$ or the gamma function to which $g(z)$ reduces after the above transformation, it is more illuminating to find the basic descriptive parameters of the exact distribution of z and from them obtain a normal approximation. Such a procedure will show how rapidly the distribution of z approaches normality with increasing N. By using the recurrence formula connecting M_{k+1} and M_k, which can be found directly from the ratio of these two moments, and expressing the necessary moments in

[3] K. Pearson, Tables of the Incomplete Gamma Function, Biometric Laboratory (1922), Univ. of London.

terms of M_1, it can be shown that these basic descriptive parameters are expressible in expanded form as follows:

$$m = \phi\left[1 - \frac{n(n+1)}{2N} + \frac{n(n+1)(n-1)(3n+2)}{24\,N^2} + \cdots\right]$$

$$\sigma^2 = \phi^2\left[\frac{2n}{N} - \frac{n(2n^2 - n + 1)}{N^2} + \cdots\right]$$

$$\beta_1 = \frac{2(3n-1)^2}{nN}\left[1 - \frac{(n+1)(5n-3)}{2(3n-1)N} + \cdots\right]$$

$$\beta_2 = 3\left[1 + \frac{4(3n-1)(4n-1)}{3nN} + \cdots\right].$$

These values suggest that

$$w = \sqrt{\frac{N}{2n}}\left[\frac{z}{\phi} - 1\right] \tag{4}$$

will likely be distributed approximately normally with zero mean and unit variance. As a matter of fact, by using the second limit theorem of probability,[4] it can be shown that the distribution of w approaches normality as N increases indefinitely. Hence, for samples in which N is large compared with n^2, it will be sufficient to compare the value of w arising from a sample $z = Z$ with its variance of unity if a test of significance is desired. A better general approximation could have been obtained by centering the curve at $\phi\left[1 - \frac{n(n+1)}{2N}\right]$ rather than at ϕ; however, since there is positive skewness and the true mean lies between these two values, there might arise some exaggeration in a significance test in doing so because the accuracy of such a test depends upon the accuracy of the approximation in the right hand tail of the curve.

why?

Inspection of (3) and (4) shows that the only population parameter upon which these approximation distributions depend is ϕ. There are no assumptions necessary about the population means, or variances, or covariances, except in so far as they may be related when the value of ϕ is postulated. This means that either (3) or (4) enables one to test whether or not it is reasonable to assume that the sample variance $z = Z$ arose in random sampling from some normal population with ϕ equal to the postulated value.

3. Population Assumptions

Consider the set of variables $u_1, u_2, \cdots, u_n$ distributed according to the normal law

$$P(u_1, u_2, \cdots, u_n) = K_1 e^{-\sum_1^n b_{ij} u_i u_j} \tag{5}$$

[4] See, for example, Frechet and Shohat, A Proof of the Generalized Second Limit Theorem in the Theory of Probability, Transactions of the American Mathematical Society, Vol. 33, (1931), p. 533.

and the set of variables $v_1, v_2, \cdots, v_n$ distributed according to the normal law

$$P(v_1, v_2, \cdots, v_n) = K_2 e^{-\sum_1^n c_i v_i^2} \tag{6}$$

where the v's are uncorrelated with the u's and with each other. The joint distribution of the u's and v's is expressed by

$$P(u_1, \cdots, v_n) = K_3 e^{-\sum_1^n b_{ij} u_i u_j - \sum_1^n c_i v_i^2}. \tag{7}$$

Upon writing down the determinant of the coefficients of these $2n$ variables, it will become evident that any one of its principal minors of any order can be expressed as the product of a principal minor of $| b_{ij} |$ with a principal minor of $| c_i |$. Since the distributions (5) and (6) are normal, the determinants $| b_{ij} |$ and $| c_i |$ are positive definite; consequently the determinant of the coefficients in (7) must also be positive definite.

Now consider the orthogonal transformation

$$y_i = \frac{u_i + v_i}{\sqrt{2}}, \qquad i = 1, 2, \cdots, n$$

$$y_i = \frac{u_i - v_i}{\sqrt{2}}, \qquad i = n + 1, \cdots, 2n.$$

Since the determinant of the coefficients in (7) is invariant under an orthogonal transformation, the resulting distribution of the y's may be expressed by

$$P(y_1, y_2, \cdots, y_{2n}) = K_4 e^{-\sum_1^{2n} d_{ij} y_i y_j} \tag{8}$$

where $| d_{ij} |$ is positive definite.

In order to obtain the distribution of the variables $y_1, y_2, \cdots, y_n$, it is necessary to integrate (8) with respect to the variables $y_{n+1}, \cdots, y_{2n}$ over their range of values. If this integration is performed after the quadratic form in the exponent of (8) has been expressed as a sum of squares[5] with coefficients which are the ratios of principal minors of $| d_{ij} |$, it will be clear that the integration leaves a quadratic form in the exponent which is also positive definite. Hence after the transformation $x_i = \sqrt{2} y_i (i = 1, 2, \cdots, n)$ the distribution function of the variables $x_i = u_i + v_i (i = 1, 2, \cdots, n)$ must be normal and may be expressed by (1). Thus it has been shown that if the true parts u_i of the variables x_i are normally distributed without error and if the error parts v_i are normally distributed but are uncorrelated with the u_i and with each other, then the variables x_i possess a normal distribution. The advantage of

[5] See, for example, Risser and Traynard, Les Principes de la Statistique Mathematique, 1933, p. 225.

this formulation will become evident when the parameter ϕ is expressed in terms of the parameters of (5) and (6).

Since the v's are uncorrelated with the u's and with each other, the variance σ_i^2 of x_i is the sum of the variances of u_i and v_i, while the correlation ρ_{ij} between x_i and x_j may be expressed in terms of the correlation ρ'_{ij} between u_i and u_j and the variances u_i^2, u_j^2, ν_i^2, ν_j^2 of u_i, u_j, v_i, v_j respectively. These relationships are

$$\sigma_i^2 = \mu_i^2 + \nu_i^2, \quad \text{and} \quad \rho_{ij} = \frac{\rho'_{ij}}{\sqrt{(1 + \nu_i^2/\mu_i^2)(1 + \nu_j^2/\mu_j^2)}} \qquad (i \neq j). \tag{9}$$

For simplicity of notation let $\lambda_i = \nu_i^2/\mu_i^2$. Now it is well known[6] that ϕ can be expressed in the form

$$\phi = \sigma_1^2\sigma_2^2 \cdots \sigma_n^2 \mid \rho_{ij} \mid .$$

If the values from (9) are inserted in $\mid \rho_{ij} \mid$ and if the resulting denominators of elements are factored out, ϕ will assume the form

$$\phi = \frac{\sigma_1^2\sigma_2^2 \cdots \sigma_n^2 B}{(1 + \lambda_1) \cdots (1 + \lambda_n)}$$

where

$$B = \begin{vmatrix} 1 + \lambda_1 & \rho'_{12} \cdots & \rho'_{1n} \\ \rho'_{12} & & \vdots \\ \vdots & & \vdots \\ \rho'_{1n} & \cdots\cdots & 1 + \lambda_n \end{vmatrix}.$$

Following the methods of confluence analysis,[7] B can be expressed as follows:

$$B = R + \sum_{\alpha=1}^{n} \lambda_\alpha R_{)\alpha(} + \sum_{\alpha<\beta}^{n} \lambda_\alpha\lambda_\beta R_{)\alpha\beta(} + \cdots + \lambda_1\lambda_2 \cdots \lambda_n$$

where $R = \mid \rho'_{ij} \mid$, $R_{)\alpha(}$ is the principal minor of R obtained by deleting row and column α, etc. R is the true correlation determinant whose rank it is the object of this paper to test. If R is assumed to be of rank $n - t$, then all principal minors containing more than $n - t$ rows vanish and B reduces to

$$B = \sum_{\alpha_1 < \cdots < \alpha_t}^{n} \lambda_{\alpha_1}\lambda_{\alpha_2} \cdots \lambda_{\alpha_t} R_{)\alpha_1\alpha_2\cdots\alpha_t(} + \cdots + \lambda_1\lambda_2 \cdots \lambda_n .$$

The tests (3) and (4) were designed to test hypothetical values of ϕ by means of the sample Z. Evidently the value of ϕ can be postulated by assigning hypothetical values to the λ's, the σ's, and the principal minors of R.

Assigning values to the λ's does not curtail the degrees of freedom in these

---

[6] S. S. Wilks, loc. cit., p. 477.

[7] Ragnar Frisch, Statistical Confluence Analysis by Means of Complete Regression Systems, Oslo, 1934.

tests because they were derived on the basis of (1) which depends only on the m's, σ's, and ρ's. The λ's do restrict the range of the ρ's, but not their degrees of freedom.

An inspection of the expression for ϕ shows that ϕ can be made to assume any desired value irregardless of the rank of R by merely assigning the σ's properly. It is therefore necessary to make some assumption regarding the σ's if the test is to serve the purpose for which it is intended. Here it will be sufficient to assume that the product of the population variances may be replaced by the product of the sample variances. This assumption will ordinarily be approximately fulfilled for the size samples for which it is legitimate to employ (3) or (4); consequently this assumption does not restrict the range of application of the test.

To postulate values of the principal minors of R beyond postulating the rank of R would introduce hypotheses and restrictions which are irrelevant to the fundamental purpose of the test. This difficulty will be avoided by replacing all non-vanishing minors of R by their upper bounds of unity. Since this will overestimate the value of B, and hence of ϕ, the usual significance level of .05 may be considered as decisive. Let the value of B when unity is inserted for all non-vanishing principal minors be denoted by D. Then

$$D = \sum_{\alpha_1 < \cdots < \alpha_t}^{n} \lambda_{\alpha_1} \lambda_{\alpha_2} \cdots \lambda_{\alpha_t} + \cdots + \lambda_1 \lambda_2 \cdots \lambda_n. \tag{10}$$

Since

$$\prod_1^n (1 + \lambda_i) = 1 + \sum_{\alpha=1}^{n} \lambda_\alpha + \sum_{\alpha_1 < \alpha_2}^{n} \lambda_{\alpha_1} \lambda_{\alpha_2} + \cdots + \lambda_1 \lambda_2 \cdots \lambda_n$$

it will often be convenient to write D in the form

$$D = \prod_1^n (1 + \lambda_i) - \left\{ 1 + \sum_{\alpha=1}^{n} \lambda_\alpha + \cdots + \sum_{\alpha_1 < \cdots < \alpha_{t-1}}^{n} \lambda_{\alpha_1} \lambda_{\alpha_2} \cdots \lambda_{\alpha_{t-1}} \right\}. \tag{11}$$

As a consequence of all the above assumptions,

$$\begin{aligned} \frac{Z}{\phi} &= \frac{|a_{ij}|}{\phi} = \frac{(1 + \lambda_1) \cdots (1 + \lambda_n) \, | r_{ij} |}{B} \\ &\geqq \frac{(1 + \lambda_1) \cdots (1 + \lambda_n) \, | r_{ij} |}{D} \end{aligned} \tag{12}$$

where $| r_{ij} |$ is the sample correlation determinant.

All the essential material for testing the rank of the true correlation matrix is contained in (3), (4), (11), and (12). In summary, the hypothesis to be tested and the procedure to follow in performing the test are as follows.

The population of n variables from which the sample is supposed drawn is assumed to be such that (a) the true parts of the variables are normally distributed, (b) the error parts are normally distributed but are uncorrelated with the true parts and with each other, (c) the product of the variances may be replaced by the product of the sample variances, (d) the values of the λ's

are postulated as judged by the accuracy in measurement of the variables, and (e) the rank of the true correlation matrix is $n - t$.

Given the value $| r_{ij} |$ of the sample correlation determinant, a lower bound for the value of Z/ϕ is calculated from (11) and (12). This lower bound is inserted in either (3) or (4), depending on the size of the sample. If (3) is used and if $P \leqq .05$, or if (4) is used and $w \geqq 2$, one may conclude, as judged by the sample variance, that it is very unlikely that the sample was drawn in random sampling from the population specified above. If one has reason to believe that the variables are sensibly normal as indicated above and that the postulated values of the λ's are quite accurate, then the test shows quite definitely that the postulated rank of the true correlation matrix is unsubstantiated by the sample, and therefore a higher rank should be tested until a non-significant value is obtained. Because a lower bound rather than the value of Z/ϕ is used, the test can be used on minimum ranks only, and hence a value of $Z < \phi$ will not yield a test of significance. However, the test does handle the problem for which it was designed and which is of fundamental interest, and that is to see whether or not one is justified in assuming that a sample represents only a certain minimum number of components.

4. Applications

(a) Hotelling[8] has used an example taken from other sources to illustrate his test on components. In order to compare results, this same example will be treated here under the assumptions outlined above. In this example the reliability coefficients are given. From the definition of a reliability coefficient r_i, it follows at once that $r_i = \dfrac{1}{1 + \lambda_i}$. The population values of the λ's will be set equal to the values obtained from these sample reliability coefficients. The data for this problem are

$$|r_{ij}| = .235,\ N = 140,\ n = 4,\ \lambda_1 = .087,\ \lambda_2 = .119,\ \lambda_3 = .101,\ \lambda_4 = .773.$$

Assume that the true correlation matrix in the population is of rank two, that is, that two components are sufficient to describe the results on these tests. Since N is large compared with n^2, it will be sufficient to use (4). The values of (11), (12), and (4) are found to be

$$D = \prod_1^4 (1 + \lambda_i) - \left\{1 + \sum_1^4 \lambda_\alpha\right\} = .294$$

$$\frac{Z}{\phi} \geqq \frac{\prod (1 + \lambda_i) \, | r_{ij} |}{D} = 1.90$$

$$w \geqq \sqrt{\frac{140}{8}}\,[1.90 - 1] = 3.76$$

[8] Loc. cit., p. 16.

Since the standard deviation of w is unity, this value demonstrates clearly that the hypothesis of only two components is untenable as judged by the sample correlation determinant. If one assumes three components, the test will be found to yield a non-significant value. Hence it may be concluded that under the hypotheses on which the test is based, the sample does not justify the assumption of less than three components. Hotelling's test indicated the necessity for two components but was uncertain about the third, the decision resting upon a variate value of 1.31 as against a standard deviation of unity.

(b) Thurstone, in his "Vectors of Mind," considers an example taken from a series of fifteen psychological tests. After applying his centroid method to the data, he inspects his results and concludes that four components are sufficient to account for everything except random errors. It is impossible to test his conclusions explicitly as above because the size of the sample is not given and the reliability coefficients are not known. Nevertheless, if it is legitimate to assume that the sample is sufficiently large to justify the use of this test, interesting conclusions can be obtained on the assumption that only four components are needed.

Suppose that $\lambda_i = \frac{1}{2}$, which implies that the variance of error is half as large as the true sampling variance for each variable. Here (10) is more convenient than (11) for computing the value of D. The values of (10) and (12) are found to be

$$D = {}_{15}C_3(\tfrac{1}{2})^{12} + {}_{15}C_2(\tfrac{1}{2})^{13} + {}_{15}C_1(\tfrac{1}{2})^{14} + (\tfrac{1}{2})^{15} = .125$$

$$\frac{Z}{\phi} \geqq \frac{|\, r_{ij} \,|}{.0003}.$$

Evidently, the value of $|\, r_{ij} \,|$ must lie in the neighborhood of .0003 if the test is not to yield a significant result which contradicts the hypothesis. However, the correlations in $|\, r_{ij} \,|$ are given to only three decimal places, and therefore a legitimate value in the neighborhood of .0003 can not be realized. It is to be noted that the postulated values of the λ's are equivalent to postulating that all reliability coefficients are equal to $\frac{2}{3}$, a value which should be considered as unusually low. It would seem reasonable to avoid using material in which the variance of error is larger than one-half the variance of random sampling, unless the variance of random sampling is exceedingly small.

CONTRIBUTIONS TO THE THEORY OF COMPARATIVE STATISTICAL ANALYSIS. I. FUNDAMENTAL THEOREMS OF COMPARATIVE ANALYSIS[1]

By William G. Madow

This is the first of several papers in which there will be presented a general approach to the statistical examination of hypotheses which are false if any of several things are true. Phenomena requiring such a statistical theory are investigated quite frequently. As examples may be cited the studies of lag correlation in time series, periodogram analysis in geophysics, factor analysis in psychology, and analysis into components in agriculture.[2]

The theorems of this paper have one purpose: to permit the reduction of the distributions by which the hypotheses are to be tested to essentially the joint distribution of the statistics which contain the information offered by the data concerning the truth or falsity of the things which will negate the hypotheses. In order to do this it has been necessary to generalize the theorêm of Poincare on the probability that at least one of several events occur.[3] As illustrations there are stated, after Theorems III, VI, and IX, generalizations of a distribution derived by Jordan, (5) page 109.[4]

In a second paper, we shall give a complete derivation of the joint distributions necessary for the applications of the analysis of variance. A reconsideration of the Schuster periodogram will be included. In other papers these results will be extended to problems arising in the theory of regression, and to problems of the distributions of medians, etc.

The fundamental theorems of comparative analysis are now obtained in such a form that they are applicable to problems in the theory of probability no matter what the distributions may be. Some special cases of these theorems[5]

---

[1] Presented to the American Mathematical Society, March 27, 1937. Research under a grant-in-aid from the Carnegie Corporation of New York.

[2] Naturally these techniques are also useful in other branches of science then those in which they were first applied. It should be noted that by analysis into components we here refer to the work of Fisher, (2), chapter 6.

[3] See, Poincaré, (7), page 60. This theorem is attributed to Poincaré by Jordan, (5), and Fréchet, (3).

[4] This distribution states the probability that in r trials of an experiment which has exactly n possible results, these results being mutually exclusive, each of the possible results occurs at least once. Jordan's derivation has been simplified by Fréchet, (3), page 12.

[5] The theorems are, of course, part of the theory of measure and integration.

have been used in connection with the derivation of distributions of positional statistics such as the k^{th} in order of N elements,[6] and others.

Let Ω be a collection of elements x, and let Δ be a set of subsets of Ω. Then, the axioms which the elements of Δ are to satisfy are[7]

I. Δ is a field;[8]

II. $\Omega \epsilon \Delta$;

III. To every $A \epsilon \Delta$ there is ordered a non-negative real number $P(A)$;

IV. $P(\Omega) = 1$;

V. If $A \epsilon \Delta$ and $B \epsilon \Delta$, and $AB = 0$, then $P(A + B) = P(A) + P(B)$.

We shall regard Ω as the set of possible results of an experiment ϵ. By events we shall mean elements of Δ. The complement $\bar{A}$ of A with respect to Ω will be an element of Δ if A is an element of Δ. $\bar{A}$ consists of all elements of Ω which are not elements of A and hence is the event which occurs if and only if A does not occur.[9]

Let the subsets of Ω

$$(1) \qquad E_1, E_2, \cdots, E_k$$

be elements of Δ. Then, if $\alpha_1, \alpha_2, \cdots, \alpha_k$ is a permutation of $1, 2, \cdots, k$, the set

$$(2) \qquad E_{\alpha_1} E_{\alpha_2} \cdots E_{\alpha_j} \bar{E}_{\alpha_{j+1}} \cdots \bar{E}_{\alpha_k}$$

is an element of Δ and is the event which occurs whenever all the events $E_{\alpha_1}, E_{\alpha_2}, \cdots, E_{\alpha_j}$ occur, while none of the events $E_{\alpha_{j+1}}, E_{\alpha_{j+2}}, \cdots, E_{\alpha_k}$ occur.

The events (1) are said to be independent if and only if

$$(3) \qquad P(E_{\alpha_1} \cdots E_{\alpha_j} \bar{E}_{\alpha_{j+1}} \cdots \bar{E}_{\alpha_k}) = \prod_{\nu=1}^{j} P(E_{\alpha_\nu}) \cdot \prod_{\nu=j+1}^{k} P(\bar{E}_{\alpha_\nu})$$

for all selections of the sets (1) and their complements.[10]

Theorem I. The probability that the first j of the k events (1) *occur, while the remaining $k - j$ events do not occur, is*

---

[6] See, for example, Gumbel, (4). It is noted that Theorems I, II, and III are stated by Arne Fisher, (1), page 42, who assumes, however, that the events are independent.

[7] These axioms are stated by Kolmogoroff, (6), page 2.

[8] A set of sets is a field if the fact that A and B are elements of the set implies that $A + B$, AB, and $A - AB$ are also elements of the set.

[9] The event A will be said to have occurred if the result of the performance of the experiment E is an element of A.

[10] See Kolmogoroff, (6), page 9 for a discussion of various equivalent definitions of independence.

$$(4)\quad P(E_1 \cdots E_j\bar{E}_{j+1} \cdots \bar{E}_k) = \sum_{\nu=0}^{k-j} (-1)^\nu \sum_{\substack{\alpha_1,\cdots,\alpha_\nu=j+1\\ \alpha_1<\alpha_2<\cdots<\alpha_\nu}}^{k} P(E_1 \cdots E_j E_{\alpha_1} \cdots E_{\alpha_\nu}).$$[11]

Proof. Let $k = j + 1$. Then it follows from Axiom V that

$$(5)\qquad P(E_1E_2 \cdots E_j) = P(E_1E_2 \cdots E_jE_{j+1}) + P(E_1E_2 \cdots E_j\bar{E}_{j+1}).$$

Hence the theorem is true for $k = j + 1$ and any $j > 0$. Let the theorem be true for $k = j, j + 1, \cdots, k - 1$. From Axiom V it follows that

$$(6)\quad P(E_1 \cdots E_j\bar{E}_{j+1} \cdots \bar{E}_k)$$
$$= P(E_1 \cdots E_j\bar{E}_{j+1} \cdots \bar{E}_{k-1}) - P(E_1 \cdots E_j\bar{E}_{j+1} \cdots \bar{E}_{k-1}E_k).$$

Substituting from (4) the theorem is proved.

Let $n \geq n_1 + \cdots + n_t$, $n_i \geq 0$ $(i = 1, \cdots, t)$; and let

$$\frac{n!}{n_1!\, n_2! \cdots n_t!\, (n - n_1 - \cdots - n_t)!} \equiv (n; n_1, n_2, \cdots, n_t).$$

COROLLARY. If, for each value of ν, $(\nu = 1, 2, \cdots, k - j)$, the $(k - j; \nu)$ terms

$$P(E_1 \cdots E_jE_{\alpha_1} \cdots E_{\alpha_\nu})$$

which can be obtained by selecting $\alpha_1, \alpha_2, \cdots, a_\nu$ without repetition from $j + 1, j + 2, \cdots, k$, are all equal, then

$$(7)\qquad P(E_1 \cdots E_j\bar{E}_{j+1} \cdots \bar{E}_k) = \sum_{\nu=0}^{k-j} (-1)^\nu(k - j; \nu)P(E_1 \cdots E_{j+\nu}).$$

Let

$$(8)\qquad S(\nu) = \sum_{\substack{\alpha_1,\cdots,\alpha_\nu=1\\ \alpha_1<\cdots<\alpha_\nu}}^{k} P(E_{\alpha_1}E_{\alpha_2} \cdots E_{\alpha_\nu})$$

where the summation extends over the $(k; \nu)$ terms

$$(9)\qquad P(E_{\alpha_1}E_{\alpha_2} \cdots E_{\alpha_\nu})$$

which can be obtained by selecting ν of the k events (1) without repetition. If all the terms (9) which can be obtained by selecting ν of the k events (1) without repetition are equal, then

$$(10)\qquad S(\nu) = (k; \nu)P(E_1 \cdots E_\nu).$$

[11] By definition

$$\sum_{\nu=0}^{k-j} (-1)^\nu \sum_{\substack{\alpha_1,\cdots,\alpha_\nu=j+1\\ \alpha_1<\cdots<\alpha_\nu}}^{k} P(E_1 \cdots E_j\bar{E}_{j+1} \cdots \bar{E}_{\alpha_\nu})$$
$$= P(E_1 \cdots E_j) + \sum_{\nu=1}^{k-j} (-1)^\nu \sum_{\substack{\alpha_1,\cdots,\alpha_\nu=j+1\\ \alpha_1<\cdots<\alpha_\nu}}^{k} P(E_1 \cdots E_jE_{\alpha_1} \cdots E_{\alpha_\nu}).$$

Theorem II. The probability that exactly j of the k events (1) *occur is*

$$P_{(j)} = \sum_{\nu=0}^{k-j} (-1)^{\nu}(j + \nu; \nu)S(j + \nu). \tag{11}$$

Proof. If $A_{(j)}$ is the subset of Ω defined by the requirement that exactly j of the events (1) occur, then $A_{(j)}$ is the sum of $(k; j)$ disjunct sets:

$$A_{(j)} = \sum_{\alpha_1, \cdots, \alpha_j = 1}^{k} E_{\alpha_1} \cdots E_{\alpha_j} \bar{E}_{\alpha_{j+1}} \cdots \bar{E}_{\alpha_k}, \tag{12}$$

where $\alpha_{j+1}, \cdots, \alpha_k$ have those of the values $1, \cdots, k$ which remain after the selection of $\alpha_1, \cdots, \alpha_j$. By Axiom V we may replace A by P in (12). Upon substituting from (4) we note that the resulting terms of (12) which depend on the same number ν, $\nu = j, \cdots, k$, of events have the same sign, that all $S(\nu)$, $\nu = j, \cdots, k$, occur, that no term depending on fewer than j events occurs, and that any particular $P(E_{\alpha_1}E_{\alpha_2} \cdots E_{\alpha_{j+t}})$ will occur in those of the terms of (12) the j occurring events of which are a subset of $E_{\alpha_1}, E_{\alpha_2}, \cdots, E_{\alpha_{j+t}}$ and will occur in no other term of (12). Hence the coefficient of $S(j + t)$ in (11) is $(-1)^t \, (j + t; t)$. This completes the proof of the theorem.

Corollary. If (10) is true for $\nu = j, \cdots, k$, then

$$P_{(j)} = \sum_{\nu=0}^{k-j} (-1)^{\nu} (k; j, \nu) \, P(E_1E_2 \cdots E_{j+\nu}). \tag{13}$$

Theorem III. The probability that at least j of the k events (1) *occur is*

$$P^{(j)} = \sum_{\nu=0}^{k-j} (-1)^{\nu} (j + \nu - 1; \nu) \, S(j + \nu). \tag{14}$$

Proof. If $A^{(j)}$ is the subset of Ω defined by the requirement that at least j of the events (1) occur, then $A^{(j)}$ is the sum of $k - j + 1$ disjunct sets:

$$A^{(j)} = A_{(j)} + A_{(j+1)} + \cdots + A_{(k)}. \tag{15}$$

By Axiom V we may replace A by P in (15). Substituting from (11)

$$P^{(j)} = \sum_{\nu=0}^{k-j} c_{\nu} \, S(j + \nu), \tag{16}$$

where

$$c_{\nu} = (j + \nu; j + \nu) - (j + \nu; 1) + \cdots + (-1)^{\nu}(j + \nu; \nu), \quad (\nu = 0, \cdots, k - j).$$

It is easy to prove that

$$(-1)^{\nu}(j + \nu - 1; \nu) = \sum_{\mu=0}^{\nu} (-1)^{\nu-\mu}(j + \nu; j + \mu). \tag{17}$$

Corollary. If (10) is true for $\nu = j, \cdots, k$, then

$$P^{(j)} = \sum_{\nu=0}^{k-j} (-1)^{\nu}(j + \nu - 1; \nu)(k; j + \nu)P(E_1E_2 \cdots E_{j+\nu}). \tag{18}$$

To provide examples illustrating these theorems let us consider r experiments

$$E^{(1)}, E^{(2)}, \cdots, E^{(r)} \tag{19}$$

Let $E^{(i)}$ have k mutually exclusive outcomes

$$O_1^{(i)}, O_2^{(i)}, \cdots, O_k^{(i)}. \tag{20}$$

Then, it is easy to define the spaces $\Omega^{(i)}$, $\Delta^{(i)}$ the probability function $P_i(E^{(i)})$, the combinatory product

$$\Omega = \Omega^{(1)} \times \Omega^{(2)} \times \cdots \times \Omega^{(r)},$$

the set Δ and the probability function $P(E)$ so that Axioms I, $\cdots$, V are satisfied and hence Theorems I, II, and III are valid.

We shall assume that the experiments (19) are independent.

Let

$$\bar{O}_j \qquad (j = 1, \cdots, k)$$

be the event which occurs when neither $O_j^{(1)}$ nor $O_j^{(2)}$ nor $\cdots$ nor $O_j^{(r)}$ occur. Then O_j occurs if upon performance of the experiments (19) at least one of $O_j^{(1)}, O_j^{(2)}, \cdots, O_j^{(r)}$ occur.

It is an immediate result of the definition of independence that

$$P(\bar{O}_{\alpha_1}\bar{O}_{\alpha_2}\cdots\bar{O}_{\alpha_j}) = \prod_{i=1}^{k}\{1 - P(O_{\alpha_1}^{(i)}) - \cdots - P(O_{\alpha_j}^{(i)})\}. \tag{21}$$

From Theorem I, the probability that $O_1, O_2, \cdots, O_j$ each occur while not one of $O_{j+1}, O_{j+2}, \cdots, O_k$ occurs is

$$\begin{aligned} P(O_1\cdots O_j\bar{O}_{j+1}\cdots\bar{O}_k) &= \sum_{\nu=0}^{j}(-1)^{\nu}\sum_{\substack{\alpha_1,\cdots,\alpha_\nu=1\\ \alpha_1<\cdots<\alpha_\nu}}^{j} \\ &\prod_{i=1}^{r}\{1 - P(O_{j+1}^{(i)}) - \cdots - P(O_k^{(i)}) - P(O_{\alpha_1}^{(i)}) - \cdots - P(O_{\alpha_\nu}^{(i)})\}. \end{aligned} \tag{22}$$

From Theorem II, the probability that exactly j of $O_1, O_2, \cdots, O_k$ occur is

$$P_{(j)} = \sum_{\nu=0}^{j}(-1)^{\nu}(k - j + \nu; \nu)S(k - j + \nu), \tag{23}$$

where

$$S(k - j + \nu) = \sum_{\substack{\alpha_1,\alpha_2,\cdots,\alpha_{k-j+\nu}=1\\ \alpha_1<\alpha_2<\cdots<\alpha_{k-j+\nu}}}^{k}\prod_{i=1}^{r}\{1 - P(O_{\alpha_1}^{(i)}) - \cdots - P(O_{\alpha_{k-j+\nu}}^{(i)})\}.$$

Since the probability that at least j of $O_1, O_2, \cdots, O_k$ occur is equal to 1 minus the probability that at least $k - j + 1$ of $\bar{O}_1, \bar{O}_2, \cdots, \bar{O}_k$ occur,[12] it follows at once from Theorem III that

$$\begin{aligned} &P\{\text{at least } j \text{ of } O_1, \cdots, O_k \text{ occur}\} = \\ &\qquad 1 - \sum_{\nu=0}^{j-1}(-1)^{\nu}(k - j + \nu; \nu)S(k - j + \nu + 1). \end{aligned} \tag{24}$$

[12] There are, of course, other ways of computing these probabilities.

The case treated by Fréchet and Jordan is that which occurs when we assume $P(O_t^{(h)}) = P(O_t^{(i)})$, $(t = 1, \cdots, k)$, $(i, h = 1, \cdots, r)$ and in (24) let $j = 1$.

It is not difficult to obtain further generalizations of Jordan's distribution by defining events which occur if and only if fewer than j' of r events occur and then proceeding as above.

Certain useful generalizations of Theorems I, II, and III will now be derived.

Let the subsets of Ω

$$E_1^{(s)}, E_2^{(s)}, \cdots, E_{k^{(s)}}^{(s)} \qquad (s = 1, \cdots, p) \tag{25}$$

be elements of Δ, and let $N = k^{(1)} + k^{(2)} + \cdots + k^{(p)}$.

Let $j^{(s)} \leq k^{(s)}$, $(s = 1, \cdots, p)$; and let

$$Q^{(t)} = \prod_{s=1}^{t} \prod_{i=1}^{j^{(s)}} E_i^{(s)} \qquad (t = 1, \cdots, p), \tag{26}$$

Let

$$Q^{(t)'} = \prod_{s=1}^{t} \prod_{i=j^{(s)}+1}^{k^{(s)}} \bar{E}_i^{(s)} \qquad (t = 1, \cdots, p). \tag{27}$$

Furthermore, let for each value of s, $(s = h, \cdots, p)$, the $(k^{(s)} - j^{(s)}; \nu^{(s)})$ possible distinct selections of $\nu^{(s)}$ of the $k^{(s)} - j^{(s)}$ sets

$$E_{j^{(s)}+1}^{(s)}, E_{j^{(s)}+2}^{(s)}, \cdots, E_{k^{(s)}}^{(s)} \tag{28}$$

be arranged in some order, and, if the intersection of the $\nu^{(s)}$ sets of the i_s^{th} selection be denoted by

$$q^{i_s}(\nu^{(s)}) \qquad (s = h, \cdots, p), \quad (i_s = 1, 2, \cdots, (k^{(s)} - j^{(s)}; \nu^{(s)})), \tag{29}$$

let

$$q^{i_h \cdots i_p}(\nu^{(h)}, \cdots, \nu^{(p)}) = \prod_{s=h}^{p} q^{i_s}(\nu^{(s)}). \tag{30}$$

There are $\prod_{s=h}^{p} (k^{(s)} - j^{(s)}; \nu^{(s)})$ sets (30), for each value of h, $(h = 1, \cdots, p)$, and any set of fixed values of $\nu^{(h)}, \cdots, \nu^{(p)}$.

Let for each value of s, $(s = h, \cdots, p)$ the $(k^{(s)}; \nu^{(s)})$ possible distinct selections of $\nu^{(s)}$ of the $k^{(s)}$ sets

$$E_i^{(s)}, \qquad (i = 1, \cdots, k^{(s)}), \tag{31}$$

be arranged in some order, and if the intersection of the sets of the i_s^{th} selection be denoted by

$$\dot{q}^{i_s}(\nu^{(s)}) \tag{32}$$

let

$$\dot{q}^{i_h \cdots i_p}(\nu^{(h)}, \cdots, \nu^{(p)}) = \prod_{s=h}^{p} \dot{q}^{i_s}(\nu^{(s)}). \tag{33}$$

There are $\prod_{s=h}^{p} (k^{(s)}; \nu^{(s)})$ sets (33), for each value of h, $(h = 1, \cdots, p)$, and any set of fixed values of $\nu^{(h)}, \cdots, \nu^{(p)}$.

It is clear that the various sets that have been defined are elements of Δ. The fact that the sets are the events which occur if and only if certain sets of events occur is also too obvious to require further comment.

Theorem IV. The probability that of the N events (25) *the first* $j^{(s)}$ *of superscript s occur and the remaining* $k^{(s)}$ *of superscript s do not occur,* $s = 1, \cdots, p$, *is*

$$(34)\quad P(Q^{(p)}Q^{(p)\prime}) = \sum_{\nu^{(1)}=0}^{k^{(1)}-j^{(1)}} \sum_{\nu^{(2)}=0}^{k^{(2)}-j^{(2)}} \cdots \sum_{\nu^{(p)}=0}^{k^{(p)}-j^{(p)}} (-1)^{\nu^{(1)}+\nu^{(2)}+\cdots+\nu^{(p)}} \sum_{i_1=1}^{(k^{(1)}-j^{(1)};\nu^{(1)})} \cdots \sum_{i_p=1}^{(k^{(p)}-j^{(p)};\nu^{(p)})} P[q^{i_1\cdots i_p}(\nu^{(1)} \cdots \nu^{(p)})].$$

Proof. Theorem I is a proof of Theorem IV for $p = 1$. The theorem may then be proved either by regarding it as a special case of Theorem I and collecting terms, or by induction.

COROLLARY. If, for each possible set of values of $\nu^{(1)}, \nu^{(2)}, \cdots, \nu^{(p)}$ the

$$\prod_{s=1}^{p} (k^{(s)} - j^{(s)}; \nu^{(s)})$$

terms

$$(35)\quad P[q^{i_1\cdots i_1}(\nu^{(1)}, \cdots, \nu^{(p)})]$$

are all equal, then

$$(36)\quad P(Q^{(p)}Q^{(p)\prime}) = \sum_{\nu^{(1)}=0}^{k^{(1)}-j^{(1)}} \cdots \sum_{\nu^{(p)}=0}^{k^{(p)}-j^{(p)}} (-1)^{\nu^{(1)}+\cdots+\nu^{(p)}} \prod_{s=1}^{p} (k^{(s)} - j^{(s)}; \nu^{(s)}) P[q^{1\cdots 1}(\nu^{(1)}, \cdots, \nu^{(p)})].$$

Let, for each value of h, $(h = 1, \cdots, p)$,

$$(37)\quad S(\nu^{(h)}, \nu^{(h+1)}, \cdots, \nu^{(p)}) = \sum_{i_h=1}^{(k^{(h)};\nu^{(h)})} \cdots \sum_{i_p=1}^{(k^{(p)};\nu^{(p)})} P[Q^{(h-1)}Q^{(h-1)\prime}\dot{q}^{i_h\cdots i_p}(\nu^{(h)} \cdots \nu^{(p)})].$$

It is apparent that by using (34) it is possible to obtain an expression for (37) which does not depend explicitly on $Q^{(h-1)\prime}$. In fact

$$(38)\quad S(\nu^{(h)}, \cdots, \nu^{(p)}) = \sum_{\nu^{(1)}=0}^{k^{(1)}-j^{(1)}} \cdots \sum_{\nu^{(h-1)}=0}^{k^{(h-1)}-j^{(h-1)}} (-1)^{\nu^{(1)}+\cdots+\nu^{(h-1)}} \sum_{i_1=1}^{(k^{(1)}-j^{(1)};\nu^{(1)})} \cdots \sum_{i_{h-1}=1}^{(k^{(h-1)}-j^{(h-1)};\nu^{(h-1)})} \sum_{i_h=1}^{(k^{(h)};\nu^{(h)})} \cdots \sum_{i_p=1}^{(k^{(p)};\nu^{(p)})} P[q^{i_1\cdots i_{h-1}}(\nu^{(1)}, \cdots, \nu^{(h-1)})\dot{q}^{i_h\cdots i_p}(\nu^{(h)}, \cdots, \nu^{(p)})].$$

If the different terms of (37) are all equal, then

$$S(\nu^{(h)}, \cdots, \nu^{(p)}) = \prod_{s=h}^{p} (k^{(s)}; \nu^{(s)}) P[Q^{(h-1)} Q^{(h-1)\prime} \dot{q}^{1, \cdots, 1}(\nu^{(h)}, \cdots, \nu^{(p)})]. \tag{39}$$

If the different terms of (38) are all equal, then

$$\begin{aligned} S(\nu^{(h)}, \cdots, \nu^{(p)}) = \sum_{\nu^{(1)}=0}^{k^{(1)}-j^{(1)}} \cdots \sum_{\nu^{(h-1)}=0}^{k^{(h-1)}-j^{(h-1)}} (-1)^{\nu^{(1)}+\cdots+\nu^{(h-1)}} \\ \prod_{s=1}^{h-1} (k^{(s)} - j^{(s)}; \nu^{(s)}) \prod_{s=h}^{p} (k^{(s)}; \nu^{(s)}) \\ P[q^{1 \cdots 1}(\nu^{(1)}, \cdots, \nu^{(h-1)}) \dot{q}^{1 \cdots 1}(\nu^{(h)}, \cdots, \nu^{(p)})]. \end{aligned} \tag{40}$$

Theorem V. The probability that of the N events (25) *the first* $j^{(s)}$ *of superscript* s *occur and the remaining* $k^{(s)}$ *do not occur,* $(s = 1, \cdots, h - 1)$, *and exactly* $j^{(s)}$ *events of superscript* s *occur* $(s = h, \cdots, p)$, *is*

$$\begin{aligned} P_{(j^{(h)} \cdots j^{(p)})}(Q^{(h-1)} Q^{(h-1)\prime}) = \sum_{\nu^{(h)}=0}^{k^{(h)}-j^{(h)}} \cdots \sum_{\nu^{(p)}=0}^{k^{(p)}-j^{(p)}} (-1)^{\nu^{(h)}+\cdots+\nu^{(p)}} \\ \prod_{s=h}^{p} (j^{(s)} + \nu^{(s)}; \nu^{(s)}) S(j^{(h)} + \nu^{(h)}, \cdots, j^{(p)} + \nu^{(p)}). \end{aligned} \tag{41}$$

Proof. The theorem may be proved, either by induction using Theorem II, or by obtaining disjunct sets as in Theorem II and using Theorem IV.

COROLLARY I. If (39) is true for all sets of possible values of $\nu^{(h)}, \cdots, \nu^{(p)}$ then

$$\begin{aligned} P_{(j^{(h)} \cdots j^{(p)})}(Q^{(h-1)} Q^{(h-1)\prime}) = \sum_{\nu^{(h)}=0}^{k^{(h)}-j^{(h)}} \cdots \sum_{\nu^{(p)}=0}^{k^{(p)}-j^{(p)}} (-1)^{\nu^{(h)}+\cdots+\nu^{(p)}} \\ \prod_{s=h}^{p} (k^{(s)}; j^{(s)}, \nu^{(s)}) P[Q^{(h-1)} Q^{(h-1)\prime} \dot{q}^{1 \cdots 1}(\nu^{(h)}, \cdots, \nu^{(p)}). \end{aligned} \tag{42}$$

COROLLARY II. If (40) is true for all sets of possible values of $\nu^{(1)}, \nu^{(2)}, \cdots, \nu^{(p)}$ then

$$\begin{aligned} P_{(j^{(h)} \cdots j^{(p)})}(Q^{(h-1)} Q^{(h-1)\prime}) = \sum_{\nu^{(1)}=0}^{k^{(1)}-j^{(1)}} \cdots \sum_{\nu^{(p)}=0}^{k^{(p)}-j^{(p)}} (-1)^{\nu^{(1)}+\cdots+\nu^{(p)}} \\ \prod_{s=1}^{h-1} (k^{(s)} - j^{(s)}; \nu^{(s)}) \prod_{s=h}^{p} (k^{(s)}; j^{(s)}, \nu^{(s)}) \\ P[q^{1 \cdots 1}(\nu^{(1)}, \cdots, \nu^{(h-1)}) \dot{q}^{1 \cdots 1}(\nu^{(h)} \cdots \nu^{(p)})]. \end{aligned} \tag{43}$$

Theorem VI. The probability that of the N events (25) *the first* $j^{(s)}$ *events of superscript* s *occur and the remaining* $k^{(s)}$ *do not occur,* $s = 1, \cdots, g - 1$, *exactly* $j^{(s)}$ *events of superscript* s *occur* $(s = g, \cdots, h - 1)$, *and at least* $j^{(s)}$ *events of superscript* s *occur* $(s = h, \cdots, p)$ *is*

$$P_{(j^{(g)}\cdots j^{(h-1)})}^{(j^{(h)}\cdots j^{(p)})}(Q^{(g-1)}Q^{(g-1)\prime}) = \sum_{\nu^{(g)}=0}^{k^{(g)}-j^{(g)}} \cdots \sum_{\nu^{(p)}=0}^{k^{(p)}-j^{(p)}} (-1)^{\nu^{(g)}+\cdots+\nu^{(p)}}$$

$$\prod_{s=g}^{h-1} (j^{(s)} + \nu^{(s)}; \nu^{(s)}) \prod_{s=h}^{p} (j^{(s)} + \nu^{(s)} - 1; \nu^{(s)}) \tag{44}$$

$$S(j^{(g)} + \nu^{(g)}, \cdots, j^{(p)} + \nu^{(p)}).$$

Proof. The theorem may be proved either by induction using Theorem III or by obtaining disjunct sets as in Theorem III and using Theorem V.

COROLLARY I. If (39) is true for all sets of possible values of

$$\nu^{(g)}, \nu^{(g+1)}, \cdots, \nu^{(p)}$$

then

$$P_{(j^{(g)}\cdots j^{(h-1)})}^{(j^{(h)}\cdots j^{(p)})}(Q^{(g-1)}Q^{(g-1)\prime}) = \sum_{\nu^{(g)}=0}^{k^{(g)}-j^{(g)}} \cdots \sum_{\nu^{(p)}=0}^{k^{(p)}-j^{(p)}} (-1)^{\nu^{(g)}+\cdots+\nu^{(p)}}$$

$$\prod_{s=g}^{h-1} (k^{(s)}; j^{(s)}, \nu^{(s)}) \prod_{s=h}^{p} [(j^{(s)} + \nu^{(s)} - 1; \nu^{(s)})(k^{(s)}; j^{(s)} + \nu^{(s)})] \tag{45}$$

$$P[Q^{(h-1)}Q^{(h-1)\prime}\dot{q}^{1\cdots 1}(\nu^{(g)}, \cdots, \nu^{(p)})].$$

COROLLARY II. If (40) is true for all sets of possible values of $\nu^{(1)}, \nu^{(2)}, \cdots, \nu^{(p)}$ then

$$P_{(j^{(g)}\cdots j^{(h-1)})}^{(j^{(h)}\cdots j^{(p)})}(Q^{(g-1)}Q^{(g-1)\prime}) = \sum_{\nu^{(1)}=0}^{k^{(1)}-j^{(1)}} \cdots \sum_{\nu^{(p)}=0}^{k^{(p)}-j^{(p)}} (-1)^{\nu^{(1)}+\cdots+\nu^{(p)}}$$

$$\prod_{s=1}^{g-1} (k^{(s)} - j^{(s)}; \nu^{(s)}) \prod_{s=g}^{h-1} (k^{(s)}; j^{(s)}, \nu^{(s)}) \prod_{s=h}^{p} [(j^{(s)} + \nu^{(s)} - 1; \nu^{(s)})(k^{(s)}; j^{(s)} + \nu^{(s)})] \tag{46}$$

$$P[q^{1\cdots 1}(\nu^{(1)}, \cdots, \nu^{(g-1)})\,\dot{q}^{1\cdots 1}(\nu^{(g)}, \cdots, \nu^{(p)})].$$

Let us again consider the experiments (19), and let us assume that $E^{(i)}$, $(i = 1, \cdots, r)$ has as its mutually exclusive results

$$O_{ts}^{(i)} \qquad (t = 1, \cdots, k^{(s)}); \; (s = 1, 2). \tag{47}$$

Let O_{ts} be the event which occurs if, upon performance of the experiments (19) at least one of the events $O_{ts}^{(1)}, O_{ts}^{(2)}, \cdots, O_{ts}^{(r)}$ occur, and let $\bar{O}_{ts}$ be the event which occurs if and only if O_{ts} does not occur.

We may state the probability that the event E_1, which occurs if and only if at least $j^{(1)}$ of the events O_{t1}, $(t = 1, \cdots, k^{(1)})$ occur, and the event E_2, which occurs if and only if at least $j^{(2)}$ of the events O_{t2}, $(t = 1, \cdots, k^{(2)})$ occur, both occur.

It is apparent that

$$P(E_1E_2) = 1 - P(\bar{E}_1) - P(\bar{E}_2) + P(\bar{E}_1\bar{E}_2), \tag{48}$$

where $\bar{E}_3$ is the event which occurs if and only if E_s does not occur, $(s = 1, 2)$.

From Theorem III

$$P(\bar{E}_s) = \sum_{\nu^{(s)}=0}^{j^{(s)}-1} (-1)^{\nu^{(s)}}(k^{(s)} - j^{(s)} + \nu^{(s)}; \nu^{(s)}) S^{(s)}(k^{(s)} - j^{(s)} + \nu^{(s)} + 1) \tag{49}$$

$$(s = 1, 2),$$

where

$$S^{(s)}(k^{(s)} - j^{(s)} + \nu^{(s)} + 1) = \sum_{\substack{\alpha_1, \cdots, \alpha_{k^{(s)}-j^{(s)}+\nu^{(s)}+1}=1 \\ \alpha_1 < \cdots < \alpha_{k^{(s)}-j^{(s)}+\nu^{(s)}+1}}}^{k^{(s)}} \tag{50}$$

$$\prod_{i=1}^{r} \{1 - P(O_{\alpha_1 s}^{(i)}) - \cdots - P(O^{(i)}_{\alpha_{k^{(s)}-j^{(s)}+\nu^{(s)}+1}s})\}, \qquad (s = 1, 2).$$

From Theorem VI

$$P(\bar{E}_1\bar{E}_2) = \sum_{\nu^{(1)}=0}^{j^{(1)}-1} \sum_{\nu^{(2)}=0}^{j^{(2)}-1} (-1)^{\nu^{(1)}+\nu^{(2)}} \prod_{s=1}^{2} (k^{(s)} - j^{(s)} + \nu^{(s)}; \nu^{(s)}) \tag{50}$$

$$S(k^{(1)} - j^{(1)} + \nu^{(1)} + 1, k^{(2)} - j^{(2)} + \nu^{(2)} + 1),$$

where

$$S(k^{(1)} - j^{(1)} + \nu^{(1)} + 1, k^{(2)} - j^{(2)} + \nu^{(2)} + 1) = \sum_{i_1=1}^{(k^{(1)}; j^{(1)}-\nu^{(1)}-1)} \sum_{i_2=1}^{(k^{(2)}; j^{(2)}-\nu^{(2)}-1)}$$

$$P[q^{i_1 i_2}(k^{(1)} - j^{(1)} + \nu^{(1)} + 1, k^{(2)} - j^{(2)} + \nu^{(2)} + 1)],$$

and

$$P[q^{i_1 i_2}(k^{(1)} - j^{(1)} + \nu^{(1)} + 1, k^{(2)} - j^{(2)} + \nu^{(2)} + 1)] =$$

$$\prod_{i=1}^{r} \left\{1 - \sum_{\nu=1}^{k^{(1)}-j^{(1)}+\nu^{(1)}+1} P(O^{(i)}_{\alpha_\nu 1}) - \sum_{\mu=1}^{k^{(2)}-j^{(2)}+\nu^{(2)}+1} P(O^{(i)}_{\beta_\mu 2})\right\},$$

the subscripts α_ν, $(\nu = 1, \cdots, k^{(1)} - j^{(1)} + \nu^{(1)} + 1)$, being those of the i_2^{th} selection of $k^{(1)} - j^{(1)} + \nu^{(1)} + 1$ events from $k^{(1)}$ events, and the subscripts β_μ, $(\mu = 1, \cdots, k^{(2)} - j^{(2)} + \nu^{(2)} + 1)$, being those of the i_2^{th} selection of $k^{(2)} - j^{(2)} + \nu^{(2)} + 1$ events from $k^{(2)}$ events.

The desired probability is then obtained by substituting from (49) and (50) into (48). The procedure is perfectly general, and applies directly to situations in which $p > 2$.

We shall now investigate the results obtained by requiring that the events considered satisfy a relation of implication.

Let the subsets of Ω

$$E_{1s}, E_{2s}, \cdots, E_{ks}, \qquad (s = 1, \cdots, p), \tag{51}$$

be elements of Δ, and let

$$E_{is} \subset E_{it}, \qquad (i = 1, \cdots, k), \tag{52}$$

if $s < t$.

It follows that

$$P(E_{is}E_{it}) = P(E_{is}), \qquad (i = 1, \cdots, k), \ (s < t). \tag{53}$$

Let $j_1 \leq j_2 \leq \cdots \leq j_t$ and let

$$Q_t = \prod_{s=1}^{t} \prod_{i=1}^{j_s} E_{is}, \qquad (t = 1, 2, \cdots, p). \tag{54}$$

Let $j_1 \leq j_2 \leq \cdots \leq j_t$ and let

$$Q'_t = \prod_{s=1}^{t} \prod_{i=j_s+1}^{k} \bar{E}_{is}, \qquad (t = 1, 2, \cdots, p). \tag{55}$$

From (52) and (53), it follows that

$$\begin{aligned} P(Q_t Q'_t) = P\Bigg(&\left[\prod_{s=1}^{t} \prod_{i=j_{s-1}+1}^{j_s} E_{is}\right] \\ &\left[\prod_{s=1}^{t-1} \prod_{i=j_s+1}^{j_{s+1}} \bar{E}_{is}\right] \prod_{i=j_t+1}^{k} \bar{E}_{it}\Bigg), \quad (j_0 = 0) \quad (t = 1, 2, \cdots, p). \end{aligned} \tag{56}$$

Let $j_1 \leq j_2 \leq \cdots \leq j_p$ and for each value of s, $(s = 1, \cdots, p)$, consider a selection of $j_s + \nu_s$ events of second subscript s from (51). Let the p selections thus obtained be such that

$$j_s + \nu_s \leq j_{s+1}, \quad (s = 1, 2, \cdots, p), \ (j_{p+1} = k),$$

and if E_{is} is one of the events of the selection of events of second subscript s then the fact that $t > s$ implies that E_{it} is one of the events of the selection of events of second subscript t.

From (52) and (53), the probability of the occurrence of all the events of the p selections thus obtained is a function of $j_p + \nu_p$ events, μ_s of which are of second subscript s, $(s = 1, \cdots, p)$ where

$$\mu_1 + \mu_2 + \cdots + \mu_s = j_s + \nu_s, \qquad (s = 1, \cdots, p), \tag{57}$$

and for a given set of values of $j_1, j_2, \cdots, j_p$ the μ_s and ν_s determine one another uniquely, $(s = 1, \cdots, p)$.

For a definite set of values of $j_1, \cdots, j_p$ and $\mu_1, \cdots, \mu_p$ or $j_1, \cdots, j_p$ and $\nu_1, \cdots, \nu_p$ there will be

$$(j_{s+1} - j_s; \nu_s) = (j_{s+1} - j_s; j_{s+1} - \mu_1 - \cdots - \mu_s), \quad (s = 1, \cdots, p), \ (j_{p+1} = k)$$

possible distinct selections of $j_s + \nu_s$, $(s = 1, \cdots, p)$ events of second subscript s, j_s of which are preassigned, from j_{s+1} events, $(s = 1, \cdots, p)$.

Let these selections be arranged in some order for each value of s, $s = 1, \cdots, p$, and let

$$q_{i_1 i_2 \cdots i_p}(\mu_1, \mu_2, \cdots, \mu_p) \tag{58}$$

be the event which occurs when for all values of s, $(s = 1, \cdots, p)$, the events of the i_s^{th} selection of $j_s + \nu_s$ events of second subscript s all occur.[13]

[13] It is understood that the j_s preassigned events of second subscript s are among the j_t preassigned events of second subscript t, $(t > s)$ in the events (58).

A typical event (58) is

$$q_{1\cdots 1}(\mu_1, \cdots, \mu_p) = \prod_{s=1}^{p} \prod_{i=j_{s-1}+\nu_{s-1}+1}^{j_s+\nu_s} E_{is}, \qquad (j_0 + \nu_0 = 0). \tag{59}$$

There will be, for a definite j_s events of second subscript s, $(s = 1, \cdots, p)$

$$\prod_{s=1}^{p} (j_{s+1} - j_s; \nu_s), \qquad (j_{p+1} = k), \tag{60}$$

events such as (58).

For a definite set of values of $\mu_1, \cdots, \mu_p$ there will be, for each value of s, $(s = 1, \cdots, p)$

$$(k - \mu_{s-1} - \cdots - \mu_1; \mu_s), \qquad (s = 1, 2, \cdots, p)$$

possible distinct selections of $j_s + \nu_s$ events of second subscript s, $j_{s-1} + \nu_{s-1}$ of which are preassigned from k events, $(s = 1, \cdots, p)$.

Let these selections be arranged in some order for each value of s,

$$(s = 1, \cdots, p),$$

and let

$$q_{i_1 i_2 \cdots i_p}(\mu_1, \mu_2, \cdots, \mu_p) \tag{61}$$

be the event which occurs if and only if, for all values of s the events of the i_s^{th} set of $j_s + \nu_s$ events of second subscript s all occur, $(s = 1, \cdots, p)$, and the first subscripts of the events of the i_s^{th} set of events of second subscript s are among the first subscripts of the events of all the selections of events of second subscript greater than s, $(s = 1, \cdots, p)$.

There will be

$$(k; \mu_1, \mu_2, \cdots, \mu_p) \tag{62}$$

events (61) which may thus be obtained.

Theorem VII. The probability that of the pK events (51) *the first j_s events of second subscript s occur and the remaining $k - j_s$ events do not occur, $s = 1, \cdots, p$,* is

$$\begin{aligned} P(Q_p Q_p') = &\sum_{\nu_1=0}^{j_2-j_1} \sum_{\nu_2=0}^{j_3-j_2} \cdots \sum_{\nu_p=0}^{k-j_p} (-1)^{\nu_1+\nu_2+\cdots+\nu_p} \\ &\sum_{i_1=1}^{(j_2-j_1;\nu_1)} \sum_{i_2=1}^{(j_3-j_2;\nu_2)} \cdots \sum_{i_p=0}^{(k-j_p;\nu_p)} P[q_{i_1 i_2 \cdots i_p}(\mu_1, \mu_2, \cdots, \mu_p)], \end{aligned} \tag{63}$$

where the event Q_t determines the $j_s - j_{s-1} - \nu_{s-1}$ events of second subscript s, $(s = 1, \cdots, p)$, which have as first subscripts all numbers $1, 2, \cdots, j_s$ which are not among the $j_{s-1} + \nu_{s-1}$ numbers determined by the events of lower second subscript than s which are contained in $q_{i_1 \cdots i_p}(\mu_1, \cdots, \mu_p)$.

Proof. Expand (56) by means of Theorem IV.

COROLLARY. If, for each fixed set of values of $\mu_1, \mu_2, \cdots, \mu_p$ the terms (58), in number (60), are all equal, then

$$P(Q_pQ'_p) = \sum_{\nu_1=0}^{j_2-j_1} \sum_{\nu_2=0}^{j_3-j_2} \cdots \sum_{\nu_p=0}^{k-j_p} (-1)^{\nu_1+\nu_2+\cdots+\nu_p} \prod_{s=1}^{p} (j_{s+1} - j_s; \nu_s) \tag{64}$$
$$P[q_{1\cdots 1}(\mu_1, \mu_2, \cdots, \mu_p)] \qquad (j_{p+1} = k).$$

Let

$$T(\mu_1, \mu_2, \cdots, \mu_p) = \sum_{i_1=1}^{(k;\mu_1)} \sum_{i_2=1}^{(k-\mu_1;\mu_2)} \cdots \sum_{i_p=1}^{(k-\mu_1-\cdots-\mu_{p-1};\mu_p)} \tag{65}$$
$$P[\dot{q}_{i_1 i_2 \cdots i_p}(\mu_1, \mu_2, \cdots, \mu_p)].$$

If all the terms of (65) are equal, then

$$T(\mu_1, \cdots, \mu_p) = (k; \mu_1, \mu_2, \cdots, \mu_p) P[\dot{q}_{1\cdots 1}(\mu_1, \cdots, \mu_p)]. \tag{66}$$

Theorem VIII. The probability that of the pK events (51) *exactly j_s events of second subscript s, $s = 1, \cdots, p$ occur, is*

$$P_{(j_1\cdots j_p)} = \sum_{\nu_1=0}^{j_2-j_1} \sum_{\nu_2=0}^{j_3-j_2} \cdots \sum_{\nu_p=0}^{k-j_p} (-1)^{\nu_1+\nu_2+\cdots+\nu_p} \tag{67}$$
$$\prod_{s=1}^{p} (\mu_s; j_s - \mu_1 - \cdots - \mu_{s-1}) T(\mu_1, \mu_2, \cdots, \mu_p).$$

Proof. If $A_{(j_1, \ldots, j_p)}$ is the subset of Ω determined by the requirement that exactly j_s of the events (51) occur $(s = 1, \cdots, p)$, then $A_{(j_1, \ldots, j_p)}$ is the sum of

$$(k; j_1, j_2 - j_1, j_3 - j_2, \cdots, j_p - j_{p-1})$$

disjunct sets which may be obtained by replacing P by A in (56) and forming (56) for all selections of $j_s - j_{s-1}$ occurring events from $k - j_{s-1}$ events, $(s = 1, \cdots, p)$. By Axiom V, $P_{(j_1, \ldots, j_p)}$ is the sum of the probabilities of these disjunct sets.

Substituting from (63), it is noted that all terms (61) which depend on the same μ_s, $(s = 1, \cdots, p)$, have the same sign and that all $T(\mu_1, \mu_2, \cdots, \mu_p)$ for which

$$0 \leq \nu_s \leq j_{s+1} - j_s, \qquad (s = 1, \cdots, p),$$

appear and only those appear. Furthermore any particular term (61) will occur in those of the terms (63) the $j_s - j_{s-1}$ occurring events of second subscript s, $(s = 1, \cdots, p)$, of which contain a fixed ν_{s-1} events, the remaining $j_s - j_{s-1} - \nu_{s-1}$ events being a subset of the μ_s events of second subscript s, $(s = 1, \cdots, p)$, that actually appear in the particular term (63). Hence the coefficient of $T(\mu_1, \cdots, \mu_p)$ is

$$(-1)^{\nu_1+\cdots+\nu_p} \prod_{s=1}^{p} (\mu_s; j_s - \mu_1 - \cdots - \mu_{s-1}), \qquad (\mu_0 = 0).$$

Corollary. If (66) is true for all sets of possible values of $\mu_1, \mu_2, \cdots, \mu_p$ then

$$P_{(j_1, \cdots, j_p)} = \sum_{\nu_1=0}^{j_2-j_1} \sum_{\nu_1=0}^{j_3-j_2} \cdots \sum_{\nu_p=0}^{k-j_p} (-1)^{\nu_1+\nu_2+\cdots+\nu_p}$$

$$(k; j_1, \nu_1, j_2 - j_1 - \nu_1; \nu_2, \cdots, j_p - j_{p-1} - \nu_{p-1}, \nu_p) \tag{68}$$

$$P[\dot{q}_{1,\cdots,1}(\mu_1, \mu_2, \cdots, \mu_p)].$$

Theorem IX. The probability that of the pk events (51) *at least* j_s, *but not more than* j_{g+1}, *events of second subscript s occur,* $(s = 1, \cdots, g)$, *and exactly* j_s *events of second subscript s occur,* $(s = g + 1, \cdots, p)$ *is*

$$P^{(j_1, \cdots, j_g)}_{(j_{g+1}, \cdots, j_p)} = \sum_{\theta_2=0}^{1} \sum_{\theta_3=0}^{1} \cdots \sum_{\theta_g=0}^{1} R_{(j_{g+1} \cdots j_p)}(1, \theta_2, \cdots, \theta_g), \tag{69}$$

where, if a 1 in the i^{th} position is denoted by δ_i, $(i = 2, \cdots, g)$,

$$R_{(j_{g+1}, \cdots, j_p)}(1, \delta_1, \cdots, \delta_{\gamma_1}, 0, \cdots, 0, \delta_{\gamma_2+1}, \cdots, \delta_{\gamma_3}, 0, \cdots, 0, \cdots, \delta_{\gamma_h+1}, \cdots, \delta_g)$$

$$= \sum_{\nu_p=0}^{k-j_p} \cdots \sum_{\nu_{g+1}=0}^{j_{g+2}-j_{g+1}} \sum_{\nu_g=0}^{j_{g+1}-j_g} \cdots \sum_{\nu_{\gamma_3}=j_{\gamma_4}-j_{\gamma_3}}^{j_{\gamma_4+1}-j_{\gamma_3}-1} \sum_{\nu_{\gamma_3-1}=0}^{j_{\gamma_2}-j_{\gamma_3-1}-1} \cdots \sum_{\nu_1=0}^{j_2-j_1-1} (-1)^{\nu_1+\nu_2+\cdots+\nu_p}$$

$$(j_1 + \nu_1 - 1; \nu_1) \cdots (j_{\gamma_3} + \nu_{\gamma_3} - j_{\gamma_3-1} - \nu_{\gamma_3-1} - 1; \nu_{\gamma_3}) \tag{70}$$

$$(j_{\gamma_4} + \nu_{\gamma_4} - j_{\gamma_3} - \nu_{\gamma_3} - 1; \nu_{\gamma_4}) \cdots (j_p + \nu_p - j_{p-1} - \nu_{p-1}; \nu_p)$$

$$T(j_1 + \nu_1, \cdots, j_{\gamma_3} + \nu_{\gamma_3} - j_{\gamma_3-1} - \nu_{\gamma_3-1}, 0, \cdots, 0,$$

$$j_{\gamma_4} + \nu_{\gamma_4} - j_{\gamma_3} - \nu_{\gamma_3}, \cdots, j_p + \nu_p - j_{p-1} - \nu_{p-1}).$$

Proof. We note first that there are 2^{g-1} terms in (69). Since

$$P^{(j_1, \cdots, j_g)}_{(j_{g+1}, \cdots, j_p)} = \sum_{\lambda_g=j_g}^{j_{g+1}} \cdots \sum_{\lambda_2=j_2}^{\lambda_3} \sum_{\lambda_1=j_1}^{\lambda_2} P_{(\lambda_1 \cdots \lambda_g j_{g+1} \cdots j_p)}, \tag{71}$$

the theorem may be proved by a process of repeated summation. From (67) and (71)

$$P^{(j_1)}_{(\lambda_2 \cdots \lambda_g j_{g+1} \cdots j_p)} = \sum_{\lambda_1=j_1}^{\lambda_2} \sum_{\nu_1=0}^{\lambda_2-\lambda_1} \sum_{\nu_2=0}^{\lambda_3-\lambda_2} \cdots \sum_{\nu_p=0}^{k-j_p} (-1)^{\nu_1+\nu_2+\cdots+\nu_p}$$

$$(\lambda_1 + \nu_1; \nu_1)(\lambda_2 + \nu_2 - \lambda_1 - \nu_1; \nu_2) \cdots (j_p + \nu_p - j_{p-1} - \nu_{p-1}; \nu_p) \tag{72}$$

$$T(\lambda_1 + \nu_1, \lambda_2 + \nu_2 - \lambda_1 - \nu_1, \cdots, j_p + \nu_p - j_{p-1} - \nu_{p-1}).$$

For fixed values of $\lambda_2, \lambda_3, \cdots, \lambda_g$ there will occur in (72) all terms

$$T(j_1 + \beta_1, \lambda_2 + \nu_2 - j_1 - \beta_1, \cdots, j_p + \nu_p - j_{p-1} - \nu_{p-1}), \tag{73}$$

$$(\beta_1 = 0, \cdots, \lambda_2 - j_1), \qquad (0 \leq \nu_s \leq \lambda_{s+1} - \lambda_s), \qquad (s = 2, \cdots, p),$$

$$(\lambda_{g+s} = j_{g+s}\ s = 1, \cdots, p - g),$$

and any definite term (73) will occur in all

$$P_{(j_1+\alpha, \lambda_2, \cdots, j_p)} \tag{74}$$

for which

$$0 \leq \alpha \leq \beta_1 .$$

In (74), the definite term (73) will have coefficient

$$(75) \quad \begin{aligned} &(-1)^{\beta_1-\alpha+\nu_2+ \cdots +\nu_p}(j_1 + \beta_1 ; j_1 + \alpha)(\lambda_2 + \nu_2 - j_1 - \beta_1 ; \nu_2) \\ &\quad \cdots (j_p + \nu_p - j_{p-1} - \nu_{p-1}, \nu_p), \qquad (\alpha = 0, 1, \cdots, \beta_1), \\ &\qquad\qquad\qquad\qquad\qquad\qquad (\beta_1 = 0, \cdots, \lambda_2 - j_1). \end{aligned}$$

Hence, in (72) the definite term (73), will have coefficient

$$\begin{aligned} &(-1)^{\beta_1+\nu_2+ \cdots +\nu_p}(j_1 + \beta_1 - 1; \beta_1)(\lambda_2 + \nu_2 - j_1 - \beta_1 ; \nu_2) \\ &\qquad\qquad\qquad \cdots (j_p + \nu_p - j_{p-1} - \nu_{p-1} ; \nu_p), \end{aligned}$$

and

$$(76) \qquad P^{(j_1)}_{(\lambda_2, \cdots, j_p)} = R_{(\lambda_2 \cdots j_p)}(1).$$

We now evaluate

$$(77) \qquad P^{(j_1 j_2)}_{(\lambda_3 \cdots j_p)} = \sum_{\lambda_2 = j_2}^{\lambda_3} P^{(j_1)}_{(\lambda_2 \cdots j_p)} .$$

For any fixed values of $\lambda_3, \cdots, \lambda_g$, there will occur in (77) all terms

$$(78) \quad \begin{aligned} &T(j_1 + \beta_1, j_2 + \beta_2 - j_1 - \beta_1, \lambda_3 + \nu_3 - j_2 - \beta_2, \\ &\qquad\qquad\qquad\qquad \cdots, j_p + \nu_p - j_{p-1} - \nu_{p-1}), \end{aligned}$$

for which either $0 \leq \beta_2 \leq \lambda_3 - j_2$; $0 \leq \beta_1 \leq j_2 - j_1 - 1$ or $\beta_1 = j_2 - j_1 + \gamma$, $0 \leq \gamma \leq \lambda_3 - j_2$; $0 \leq \beta_2 \leq \lambda_3 - j_2 - \gamma$.

Let $0 \leq \beta_1 \leq j_2 - j_1 - 1; 0 \leq \beta_2 \leq \lambda_3 - j_2$. Then the term (78) will occur in all

$$(79) \qquad P^{(j_1)}_{(j_2+\alpha, \lambda_3, \cdots, j_p)},$$

such that

$$0 \leq \alpha \leq \beta_2 .$$

In (79), (78) will have coefficient

$$(80) \quad \begin{aligned} &(-1)^{\beta_1+\beta_2-\alpha+\nu_3+\cdots+\nu_p}(j_1 + \beta_1 - 1; \beta_1)(j_2 + \beta_2 - j_1 - \beta_1 - 1; \beta_2 - \alpha) \\ &\quad (\lambda_3 + \nu_3 - j_2 - \beta_2 ; \nu_3) \cdots (j_p + \nu_p - j_{p-1} - \nu_{p-1} ; \nu_p). \end{aligned}$$

Hence in (77), (78) will have coefficient

$$(81) \quad \begin{aligned} &(-1)^{\beta_1+\beta_2+\nu_3+\cdots+\nu_p}(j_1 + \beta_1 - 1; \beta_1)(j_2 + \beta_2 - j_1 - \beta_1 - 1; \beta_2) \\ &(\lambda_3 + \nu_3 - j_2 - \beta_2 ; \nu_3) \cdots (j_p + \nu_p - j_{p-1} - \nu_{p-1} ; \nu_p), \\ &(\beta_1 = 0, \cdots, j_2 - j_1 - 1), \qquad (\beta_2 = 0, \cdots, \lambda_3 - j_2), \\ &\qquad (\nu_s = 0, \cdots, \lambda_{s+1} - \lambda_s), \qquad (s = 3, \cdots, p); \\ &\qquad\qquad\qquad (\lambda_{g+s} = j_{g+s}), \ (s = 1, \cdots, p - g). \end{aligned}$$

Now let $\beta_1 = j_2 - j_1 + \gamma; 0 \leq \gamma \leq \lambda_3 - j_2 ; 0 \leq \beta_2 \leq \lambda_3 - j_2 - \gamma$. Then the term (78) will occur in all terms (79) such that

$$\gamma \leq \alpha \leq \beta_2 ,$$

and in (79), (78) will have coefficient (80). Summing for α, $(\alpha = \gamma, \cdots, \beta_2)$, we obtain as the coefficient of (78) in (77)

$$0, \qquad \text{if } \beta_2 > \gamma,$$

and

$$(-1)^{\beta_1+\nu_3+\cdots+\nu_p}(j_1 + \beta_1 - 1; \beta_1)(\lambda_3 + \nu_3 - j_1 - \beta_1 ; \nu_3) \cdots (j_p + \nu_p - j_{p-1} - \nu_{p-1} ; \nu_p), \qquad \text{if } \beta_2 = \gamma.$$

Hence

$$P^{(j_1 j_2)}_{(\lambda_3 \cdots j_p)} = R_{(\lambda_3 \cdots j_p)}(1, 1) + R_{(\lambda_3 \cdots j_p)}(1, 0). \tag{82}$$

If we examine (82), we note that the result of summing with respect to λ_2 has been the replacement of (76) by two sums which are similar to (76) in that the next summation index, in this case λ_3, occurs in exactly two limits of summation. If it can be shown that the two sums which occur in (82) each result in a pair of sums after summation with respect to λ_3, or more exactly if

$$\sum_{\lambda_{s+1}=j_s+1}^{\lambda_{s+2}} R_{(\lambda_{s+1}, \cdots, j_p)}(1, \theta_2, \cdots, \theta_s) = R_{(\lambda_{s+2} \cdots j_p)}(1, \theta_2, \cdots, \theta_s, 1) + R_{(\lambda_{s+2}, \cdots, j_p)}(1, \theta_2, \cdots, \theta_s, 0) \tag{83}$$

then the proof will be completed.

Since the truth of (83) may be demonstrated in exactly the same way in which (82) has been shown to be true, the theorem is proved.

Corollary. If (66) is true for all sets of possible values of $\mu_1, \mu_2, \cdots, \mu_p$ then

$$\begin{aligned} R_{(j_g+1, \cdots, j_p)}&(1, \delta_1, \cdots \delta_{\gamma_1}, 0, \cdots, 0, \delta_{\gamma_2+1}, \cdots, \delta_{\gamma_3}, 0, \cdots 0, \cdots, \delta_{\gamma_k+1}, \cdots, \delta_g) \\ &= \sum_{\nu_p=0}^{k-j_p} \cdots \sum_{\nu_{g+1}=0}^{j_{g+2}-j_{g+1}} \sum_{\nu_g=0}^{j_{g+1}-j_g} \cdots \sum_{\nu_{\gamma_3}=j_{\gamma_4}-j_{\gamma_3}}^{j_{\gamma_4+1}-j_{\gamma_3}-1} \cdots \sum_{\nu_1=0}^{j_2-j_1-1} (-1)^{\nu_1+\cdots+\nu_p} \\ &\quad (j_1 + \nu_1 - 1; \nu_1) \cdots (j_{\gamma_3} + \nu_{\gamma_3} - j_{\gamma_3-1} - \nu_{\gamma_3-1} - 1; \nu_{\gamma_3}) \\ &\quad (j_{\gamma_4} + \nu_{\gamma_4} - j_{\gamma_3} - \nu_{\gamma_3} - 1; \nu_{\gamma_4}) \cdots (j_p + \nu_p - j_{p-1} - \nu_{p-1} ; \nu_p) \\ &\quad (k; j_1 + \nu_1, \cdots, j_{\gamma_3} + \nu_{\gamma_3} - j_{\gamma_3-1} - \nu_{\gamma_3-1}, j_{\gamma_4} \\ &\qquad + \nu_{\gamma_4} - j_{\gamma_3} - \nu_{\gamma_3}, \cdots, j_p + \nu_p - j_{p-1} - \nu_{p-1}) \\ &\quad P[\dot{q}_{1\cdots 1}(j_1 + \nu_1, \cdots, j_{\gamma_3} + \nu_{\gamma_3} - j_{\gamma_3-1} - \nu_{\gamma_3-1}, 0, \cdots, 0, \\ &\qquad j_{\gamma_4} + \nu_{\gamma_4} - j_{\gamma_3} - \nu_{\gamma_3}, \cdots, j_p + \nu_p - j_{p-1} - \nu_{p-1})]. \end{aligned} \tag{84}$$

Let us again consider the experiments (19) and let $E^{(i)}$ have as possible results

$$O_{j_s}^{(i)} \qquad (j = 1, \cdots, k), (s = 1, 2) \ (i = 1, 2, \cdots, r).$$

Let

$$O_{j_1}^{(i)} \supset O_{j_2}^{(i)} \qquad (i = 1, \cdots, r), \ (j = 1, \cdots, k),$$

i.e. $O_{j_1}^{(i)}$ occurs whenever $O_{j_2}^{(i)}$ occurs. Furthermore let the outcomes

$$O_{11}^{(i)}, O_{21}^{(i)}, \cdots, O_{k_1}^{(i)}$$

be mutually exclusive.

Let

$$\bar{O}_{js}, \qquad (s = 1, 2),$$

occur if and only if none of

$$O_{js}^{(1)}, O_{js}^{(2)}, \cdots, O_{js}^{(k)}$$

occur.

We may wish to know the probability that at least j_1 of $\bar{O}_{11}, \cdots, \bar{O}_{k1}$ and at least j_2, $j_2 \geq j_1$, of $\bar{O}_{12}, \bar{O}_{22}, \cdots, \bar{O}_{k2}$ occur.

From Theorem IX this probability is equal to

$$P^{(j_1, j_2)} = R(1, 1) + R(1, 0), \tag{85}$$

where

$$R(1, 1) = \sum_{\nu_2=0}^{k-j_2} \sum_{\nu_1=0}^{j_2-j_1-1} (-1)^{\nu_1+\nu_2}(j_1 + \nu_1 - 1; \nu_1)$$
$$(j_2 + \nu_2 - j_1 - \nu_1 - 1; \nu_2) T(j_1 + \nu_1, j_2 + \nu_2 - j_1 - \nu_1),$$

and

$$R(1, 0) = \sum_{\nu_1=j_2-j_1}^{k-j_1} (-1)^{\nu_1}(j_1 + \nu_1 - 1; \nu_1) T(j_1 + \nu_1).$$

From (63)

$$T(j_1 + \nu_1, j_2 + \nu_2 - j_1 - \nu_1) = \sum_{i_1=1}^{(k; j_1+\nu_1)} \sum_{i_2=1}^{(k-j_1-\nu_1; j_2+\nu_2-j_1-\nu_1)} P[\dot{q}_{i_1 i_2}(j_1 + \nu_1; j_2 + \nu_2 - j_1 - \nu_1)], \tag{86}$$

where, from (61)

$$\dot{q}_{i_1 i_2}(j_1 + \nu_1, j_2 + \nu_2 - j_1 - \nu_1) = \prod_{\nu=1}^{j_1+\nu_1} \bar{O}_{\alpha_\nu 1} \prod_{\nu=j_1+\nu_1+1}^{j_2+\nu_2} \bar{O}_{\alpha_\nu 2},$$

the subscripts

$$\alpha_1, \alpha_2, \cdots, \alpha_{j_1+\nu_1} \tag{87}$$

being the first subscripts of the i_1^{th} selection of $j_1 + \nu_1$ events of second subscript 1 from

$$\bar{O}_{11}, \bar{O}_{21}, \cdots, \bar{O}_{k1},$$

and the subscripts

$$\alpha_{j_1+\nu_1+1}, \alpha_{j_1+\nu_1+2}, \cdots, \alpha_{j_2+\nu_2},$$

being the first subscripts of the i_2^{th} selection of $j_2 + \nu_2$ events of second subscript 2, $j_1 + \nu_1$ of which are (87), from

$$\bar{O}_{12}, \bar{O}_{22}, \cdots, \bar{O}_{k2}.$$

It is easy to see that

$$P[\dot{q}_{i_1 i_2}(j_1 + \nu_1, j_2 + \nu_2 - j_1 - \nu_1)] = \prod_{i=1}^{r} \left\{1 - \sum_{\nu=1}^{j_1+\nu_1} P(O_{\alpha_\nu 1}^{(i)}) - \sum_{\nu=j_1+\nu_1+1}^{j_2+\nu_2} P(O_{\alpha_\nu 2}^{(i)})\right\}.$$

Furthermore

$$T(j_1 + \nu_1) = \sum_{i_1=1}^{(k; j_1+\nu_1)} P[\dot{q}_{i_1}(j_1 + \nu_1)], \tag{88}$$

where

$$P[\dot{q}_{i_1}(j_1 + \nu_1)] = \prod_{i=1}^{r} \left\{1 - \sum_{\mu=1}^{j_1+\nu_1} P(O_{\alpha_\mu 1}^{(i)})\right\}.$$

Substituting from (86) and (88) into (85) the desired probability is obtained.

It may be remarked that theorems which have the same relation to Theorems VII, VIII, and IX that Theorems IV, V, and VI have to Theorems I, II, and III may be obtained without much difficulty.

BIBLIOGRAPHY

(1) A. Fisher, The Mathematical Theory of Probabilities, The MacMillan Company, New York, 1930.
(2) R. A. Fisher, The Design of Experiments, Oliver and Boyd, London, 1935.
(3) M. Fréchet, Généralités sur les Probabilités, Gauthier-Villars, Paris, 1937.
(4) E. J. Gumbel, Les valeurs extrêmes des distributions statistiques, Annales de L'Institut Henri Poincaré, vol. 5, (1935), pp. 115–158.
(5) Ch. Jordan, Le théorème de probabilité de Poincaré généralisé au cas de plusieurs variables indépendantes, Acta Scientarum Mathematicarum, Szeged, vol. 7, (1934), pp. 103–111.
(6) A. Kolmogoroff, Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse der Mathematik, vol. 2, no. 3.
(7) H. Poincaré, Calcul des Probabilités, Gauthier-Villars, Paris, 1912.

Columbia University.

REPLY TO MR. WERTHEIMER'S PAPER

RICHMOND T. ZOCH

The attainment of rigor both in applied as well as pure mathematics is a slow process, and for this reason criticism of my paper, if constructive, is welcomed.

Properties like continuity, differentiability, and dimensionality are *local* properties, that is to say a function may be continuous or differentiable over a certain range but not outside this range, or otherwise a function may be continuous or differentiable over a given range except for singular points.

The presence of singularities in functions does not necessarily cancel their utility. Thus the function $y = \tan x$ contains points where it is discontinuous, but ordinarily it is regarded as a continuous function and the presence of these singular points seldom handicaps one when working with this function. Similarly, the function $f = \bar{x} - \frac{1}{2}\frac{\mu_3}{\mu_2}$ is a function which satisfies all four Axioms as stated in Whittaker and Robinson's book and expresses the mode of Pearson's Type III curve as a symmetric function of the measures. The fact that this function is not differentiable along the line $x_1 = x_2 = x_3 = \cdots = x_n$ will never handicap the investigator for unless the frequency distribution is clearly skew the Type III curve would not be used to represent it.

It seems that Mr. Wertheimer bases nearly all his criticisms on the tacit addition of the word "*everywhere*" to Axiom IV as stated in Whittaker and Robinson's book. The word "everywhere" is not in the statement of Axiom IV and I assumed nothing else than stated in the axiom.

If one deliberately adds the word "everywhere" to Axiom IV then nearly all my criticisms of previous writers are incorrect, unfair, and unjust. However, it does not seem that clearness and rigor in mathematics are increased by reading into an axiom a word that is not there.

Consider first the criticism in my paper which remains valid even when the word "everywhere" is added. (Schimmack uses the word "everywhere" on page 127 although Whittaker and Robinson do not.) Both Schimmack and Whittaker and Robinson proceed as at the top of page 217 of the book by the latter authors with the statement: "In this equation make $k \to 0$ then each of the quantities $\left[\frac{\partial f}{\partial x_n}\right]$ tends to a value which is independent of the x's $\cdots$."

This statement rests on the tacit assumption that the quantities $\left[\frac{\partial f}{\partial x_n}\right]$ are functions of k. Even if such were true the use of tacit assumptions in a rigorous proof is objectionable, but as a matter of fact these quantities are not functions of k. Thus the particular proof given in Whittaker and Robinson's book as

well as in Schimmack's paper is altogether lacking in rigor even when the word "everywhere" is added to Axiom IV. Both Schiaparelli's and Broggi's proofs appear to be entirely rigorous if the word "everywhere" is added to Axiom IV.

In preparing my paper I assumed that no prohibition on functions which had singular points was contained in Axiom IV. In other words, I assumed since the word "everywhere" did not appear there was no valid objection to introduce and discuss functions with singularities. The functions I introduced are everywhere *continuous* but they are not *differentiable* along the line in Euclidian n-space defined by $x_1 = x_2 = x_3 = \cdots = x_n$. They are differentiable at every other point in the space.

It seems to me since Axiom IV as stated in Whittaker and Robinson's book does not exclude functions which are not everywhere differentiable that all my criticism is fair and just, and moreover nearly all my statements are correct. Mr. Wertheimer is entirely correct in pointing out that the words "everywhere" on page 181 of my paper are contradictory. As a matter of fact the whole paragraph beginning with line 7 on page 181 appears to me, on reexamining it, to be unsatisfactory. Except for this single paragraph I believe my paper to be rigorous, but I welcome further criticism.

Mr. Wertheimer's conclusions in his paragraph number 4 are clearly erroneous. To show this, consider a function of k. As $k \to 0$ any one of three situations may arise, namely: (1) The function may become infinite, (2) the function may become indeterminate, that is it may take on any value whatever, (3) the function may approach a unique finite value independent of k. Neither Schimmack nor Whittaker and Robinson nor Mr. Wertheimer has established as a definite fact that the particular type of function here in question approaches a unique finite value independent of k as $k \to 0$. The truth of the matter is that this conclusion cannot be established because the function in question does not involve k either explicitly or implicitly.

In conclusion there are two things I wish to emphasize. First, even when the word "everywhere" is added to Axiom IV, the proof given in Whittaker and Robinson's book is faulty, but if one consults the references given there in the footnotes he will find two other proofs which are rigorous with this addition to Axiom IV. Second, the mode of a skew bell shaped Pearson Frequency Curve satisfies all four axioms as stated in Whittaker and Robinson's book, and the fact that these expressions for the mode are not differentiable along a certain line is never a handicap to the statistician.

George Washington University.

e
t.
e
n
y

k
y
t.
,

e
t,
o

e-
n-
c-
r,
er
ed
es
at
ot

en
ter
ere
ad-
re-
n's
ble